From DEPARTMENT OF LABORATORY MEDICINE,

Karolinska Institutet, Stockholm, Sweden

STUDYING THE CELL CYCLE USING SYSTEMS BIOLOGY AND HIGH CONTENT CHARACTERIZATION OF THE UBIQUITIN PROTEASOME SYSTEM

Johan Boström

Karolinska Institutet

Stockholm 2020

Published by Karolinska Institutet.
Printed by Universitetsservice US-AB

ISBN 978-91-7831-932-9

STUDYING THE CELL CYCLE USING SYSTEMS BIOLOGY AND HIGH CONTENT CHARACTERIZATION OF THE UBIQUITIN PROTEASOME SYSTEM

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Johan Boström

Principal Supervisor:
Dr Mikael Altun
Karolinska Institutet
Department of Laboratory Medicine
Division of Clinical Physiology

Co-supervisors:
Dr Michael Andäng
Uppsala University / Masaryk University
Department of IGP / SEITEC

Prof Thomas Helleday
University of Sheffield / Karolinska Institutet
Department of Oncology and Pathology
Division of Translational Medicine

Dr Eric Rullman
Karolinska Institutet
Department of Laboratory Medicine
Division of Clinical Physiology

Opponent:
Ass Prof Ida-Maria Sintorn
Uppsala University
Department of Information Technology

Examination Board:
Prof Elias Arnér
Karolinska Institutet
Dep. of Medical Biochemistry and Biophysics

Prof Johan Trygg
Umeå University
Department of Chemistry

Prof Nico Dantuma
Karolinska Institutet
Department of Cell and Molecular Biology

“You see, Momo,' he [Beppo Roadsweeper] told her one day, 'it's like this. Sometimes, when you've a very long street ahead of you, you think how terribly long it is and feel sure you'll never get it swept.'

He gazed silently into space before continuing. 'And then you start to hurry,' he went on. 'You work faster and faster, and every time you look up there seems to be just as much left to sweep as before, and you try even harder, and you panic, and in the end you're out of breath and have to stop - and still the street stretches away in front of you. That's not the way to do it.'

He pondered a while. Then he said, 'You must never think of the whole street at once, understand? You must only concentrate on the next step, the next breath, the next stroke of the broom, and the next, and the next. Nothing else.'

Again he paused for thought before adding, 'That way you enjoy your work, which is important, because then you make a good job of it. And that's how it ought to be.'

There was another long silence. At last he went on, 'And all at once, before you know it, you find you've swept the whole street clean, bit by bit. What's more, you aren't out of breath.' He nodded to himself. 'That's important, too,' he concluded.”

– Michael Ende

ABSTRACT

The cell cycle is the process through which our cells grow and divide in a carefully orchestrated and controlled manner. An uncontrolled cell cycle is a basic hallmark of cancer, in which excessive cell growth leads to adverse effects on the tissue- and organ-level. The Ubiquitin Proteasome System (UPS) is responsible for the destruction of unwanted proteins in the cell, but has additionally been identified to inhabit regulatory roles in a myriad of cellular processes. This thesis aimed to study the cell cycle and the ubiquitin proteasome system using high-throughput and high-content approaches.

The first approach aimed at observing the endogenous fluctuations of mRNA, proteins, phosphorylations, and intra-cellular compartmentalization over the cell cycle, and how these systems are regulated and coordinated throughout the cell cycle, described in **Study I** and **II**. Aside from characterizing cell cycle oscillation patterns of transcripts, proteins, phosphorylation events and subcellular localization changes, the dynamics between transcriptional and proteomic regulation was further investigated by comparing oscillation patterns of corresponding mRNA and protein pairs.

The second approach aimed to investigate how one of the largest enzyme families in the human proteome, the UPS, affects the cell cycle and responses to external and intrinsic DNA damage. This was done through a phenotypical characterization after silencing the genes comprising the family in a high-content imaging study, **Study III**. The results revealed many novel UPS genes as essential for proper progression through the cell cycle and maintenance of DNA integrity. By combining multiple reporter systems in one high-content study, correlations between cell cycle, viability and DNA Damage Response phenotypes could be performed. This revealed an increased tendency for G1/S-phase cell cycle arrests after signs of spontaneous DNA damage, and an enrichment for G2 cell cycle arrests after failure of 53bp1 recruitment to double-strand breaks.

Aside from providing data resources and system biology results regarding the interaction between different cellular process, the studies also identified specific genes and proteins in novel roles regarding these basic cellular processes. In **Study II**, the methyl-transferase protein MAT2A was discovered to change subcellular localization in synchronization with the cell cycle, possible to provide the higher source of methyl groups needed during S-phase and G2-phase. In **Study IV**, the E3 ubiquitin ligases ARIH1 and ARIH2 were investigated for effects on proliferation and growth of glioblastoma multiforme. The high-content **Study III** identified many novel UPS genes with a myriad of cell cycle and DNA damage phenotypes, among them the E3 ubiquitin adapter BTBD1. Silencing of BTBD1 incurred dramatic phenotypes on the cell cycle and DNA damage responses, and BTBD1 was further characterized and identified to be essential for proper function of the DNA topoisomerase TOP1.

Throughout these projects, in order to validate novel findings, methods to control specific gene expression levels was developed, utilizing shRNA and CRISPR/Cas9-mediated silencing as well as a flexible method of overexpression. These systems are described in **Study IV**.

The presented studies combine high-content and high-throughput approaches with novel visualization and analysis methods to distill information from complex data, both to summarize interactions between mRNA, proteins, and function, but also to identify novel regulators of basal cellular processes.

LIST OF SCIENTIFIC PAPERS

This thesis is composed of the following papers

I. **Boström J***, Sramkova Z*, Salašová A*, Johard H, Mahdessian D, Fedr R, Marks C, Medalová J, Souček K, Lundberg E, Linnarsson S, Bryja V, Sekyrova P, Altun M, Andäng M, **Comparative Cell Cycle Transcriptomics Reveals Synchronization of Developmental Transcription Factor Networks in Cancer Cells**, Plos One, 2017, 12(12)

II. Herr P*, **Boström J***, Eric Rullman E, Rudd S.G, Vesterlund M, Lehtiö J, Helleday T, Maddalo G, Altun M, **Cell Cycle Profiling Reveals Protein Oscillation, Phosphorylation, and Localization Dynamics**, Molecular and Cellular Proteomics, 2020

III. **Boström J**, Kanellis D, Valerie N.C.K, Maddalo G, Lyu J, Zhang S, Unterlass J, Rudd S.G, Sanjiv K, Marks C, Martens U, Häggblad M, Lundgren B, Elsässer S, Helleday T, Bartek J, Altun M, **Mapping the Functional Landscape of the Ubiquitin System within the Cell Cycle and DNA Repair by High Content Imaging** – *Manuscript*

IV. **Boström J**, Harisankar A, Alam S, Valerie N.C.K, Sanjiv K, Marks C, Esthad S, Bräutigam L, Walfridsson J, Altun M, **A Multifluorescent, Multipurpose Genetic Engineering System for Studying Biological Processes.** – *Manuscript*

* = *Equal Contributions*

LIST OF ABBREVIATIONS

ANOVA	Analysis of Variance
APC	Anaphase-Promoting Complex
ATCC	American Type Culture Collection
CDK	Cyclin Dependent Kinase
ChIP	Chromatin Immuno-Precipitation
CKI	CDK Inhibitory Proteins
CPM	Counts Per Million
CPT	Camptothecin
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRL	Cullin-RING-Ligase
DAPI	4′,6-diamidino-2-phenylindole
DDR	DNA Damage Response
DNA	Deoxyribonucleic Acid
DUB	Deubiquitinating enzyme
EdU	5-Ethynyl-2′-deoxyuridine
FC	Fold Change
FP	Fluorescent Protein
Fucci	Fluorescent Ubiquitination-based Cell Cycle Indicator
G1	Gap 1 Phase
G2	Gap 2 Phase
GBM	Glioblastoma Multiforme
GEO	Gene Expression Omnibus
GFP	Green Fluorescent Protein
GO	Gene Ontology
gRNA	Guide RNA
HECT	Homologous to the E6AP Caroboxyl Terminus
HGNC	HUGO Gene Name Nomenclature
HR	Homologous Recombination
MJD	Machado Joseph Disease
mRNA	Messenger RNA
NHEJ	Non-Homologous End-Joining
OUT	Ovarian Tumor
PBS	Phosphate-Buffered Saline
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PLA	Proximity Ligation Assay
PRIDE	Proteomics Identification Database
RING	Really Interesting New Gene
RNA	Ribonucleic Acid
RNF	RING-Finger-Proteins
SCF	Skp, Cullin, F-box
SQL	Structured Query Language
shRNA	Short Hairpin RNA
siRNA	Small Interfering RNA
TF	Transcription Factor
TLS	Translesion Synthesis
t-sne	T-distributed Stochastic Neighbor Embedding
TSS	Transcription Start Site
UBL	Ubiquitin-Like
UCH	Ubiquitin-Carboxyl-terminal Hydrolase
UPS	Ubiquitin-Proteasome System

LIST OF GENES AND PROTEINS

The inclusion of this list is to provide unambiguous references to genes and proteins mentioned throughout the thesis. HUGO Gene Name Nomenclatures (HGNC) are used for almost all genes, but since well-studied genes are sometimes referred to by other names, Entrez Gene IDs and Uniprot IDs are provided here.

Gene Name	NCBI Gene ID	Uniprot ID
53BP1	7158	Q12888
ARIH1	25820	Q9Y4X5
ARIH2	10425	O95376
ATM	472	Q13315
BRCA1	672	P38398
BTBD1	53339	Q9H0C5
BTBD2	55643	Q9BX70
β-TRCP	8945	B7Z3H4
CDC20	991	Q12834
CDH1	999	P12830
CDH3	1001	P22223
CDKN1B	1027	P46527
CDT1	81620	Q9H211
E2F1	1869	Q01094
FBXW7	55294	Q969H0
Geminin	51053	O75496
ISG15	9636	P05161
KAP1	10155	Q13263
MAT2A	4144	P31153
MDM2	4193	Q00987
PAX6	5080	P26367
P53	7157	P04637
PCNA	5111	P12004
RIF1	55183	Q5UIP0
RNF168	165918	Q8IYW5
RNF8	9025	O76064
SKP2	6502	Q13309
TK	7083	P04183
TOP1	7150	P11387
UBE2N	7334	P61088
UCHL5	51377	Q9Y5K5
UHRF1	29128	Q96T88
USP10	9100	Q14694
USP14	9097	P54578
USP18	11274	Q9UMW8
USP7	7874	Q93009

CONTENTS

1 INTRODUCTION

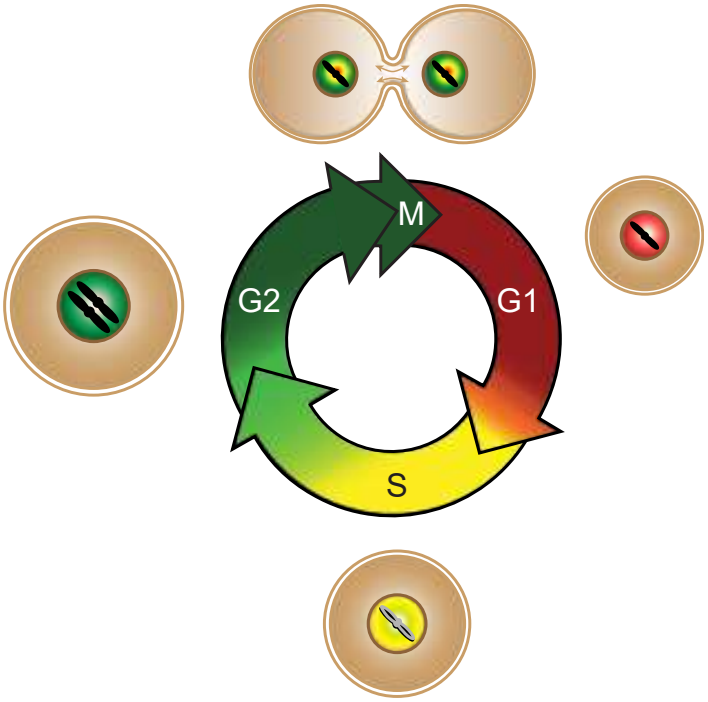

Figure 1. The Cell Cycle

1.1 STUDYING THE CELL CYCLE

The cell cycle is the regulatory system responsible for controlled replication and mitosis of cells, controlling proliferation. This proliferation is carefully orchestrated both during development and in the adult organism, and uncontrolled proliferation is a major hallmark of cancer. The broadest division of phases describes cells as going through the pre-replicative G1 phase, through the Cyclin Dependent Kinase (CDK) regulated G1/S checkpoint into the replicative S-phase. After successful replication, the G2-phase leads past the G2/M-checkpoint into mitosis, the act of separating the chromatin followed by cytokinesis, cell division (**Figure 1**). The CDK family of kinases has the most direct role on cell cycle events through the phosphorylation of effector proteins, especially to pass the cell cycle checkpoints. The activation of the CDKs is primarily controlled by the Cyclin proteins, as well as CDK inhibitory proteins (CKIs). The Cyclins oscillate over the cell cycle in a periodic fashion, giving rise to the precise events of CDK activation and subsequent activation of effector proteins (1,2).

The most common way to study the cell cycle is to synchronize cells in a specific phase using inhibitor treatment, to then remove the inhibitors and follow the cells as they develop further (3). Thymidine-block and double-thymidine block are two common ways to achieve this. The first thymidine block arrests cells throughout S-phase, and if they are then released and thymidine-blocked again almost a cell cycle later, they exhibit a clear early S-phase blockade.

To synchronize cells between G2 and mitosis, low concentrations of Nocadozole can be used. Other methods of synchronization are serum deprivation or contact inhibition, although this leads to many cells entering quiescence (4). A disadvantage with synchronization methods is that the synchronization itself deregulates the cell cycle and might cause more extreme mRNA depletions and enrichments than what would normally occur in freely dividing cells. Any chemical intervention also risks unwanted effects, which can be hard to separate from the studied biology.

Contrary to synchronization, methods of separating asynchronous cell-cycle progressing cells have become of recent interest. Some methods use simple physics. The fact that mitotic cells are rounded up and loosely adherent in a cell culture flask can be utilized to collect them with a "mitotic shake off" (5). Cells additionally grow in size over the cell cycle, which can be utilized to separate enriched portions of cell cycle phases through centrifugal elutriation (3).

Other methods utilize genetic reporters. In 2008, an interesting method was developed for cell cycle studies by the Miyawaki group at the RIKEN Institute in Japan. They developed a set of fluorescent probes, using truncated peptides from hCdt1 and hGeminin coupled to different fluorescent proteins, to visualize the SCF/APC-complex activity equilibrium, closely following the regulatory cell cycle (6). During the S phase and G2-phase, the CDT1 probe is ubiquitinated and degraded together with its fluorescent protein by the APC-Complex using the effector protein CDH1, and at mitosis the Geminin probe is degraded by the SCF-complex with SKP2, leading to a clear distinction of cell cycle phases for use both in animal models and *in vitro* studies, including live-cell imaging, flow cytometry or high-throughput imaging studies (6–9).

Two major studies in recent years have performed transcriptome-wide characterizations of cell cycle expression of proteins. The first one, by Whitfield et al, used a MicroArray-based approach to characterize transcripts from cells released from synchronization with time-points covering three full cell cycle phases (10). They then fitted sinus curve expression patterns to these as a statistical test. The second study was conducted by Dominguez et al. (2016), who used a smaller amount of time-points and a seed-matching approach to identify oscillating genes (11).

Although transcriptional regulation of protein-coding genes is a major driver of proteomic regulation, many other processes are involved in regulation of protein levels. When studying cell types and tissues, transcriptomics and proteomics can correlate (12) but mRNA is not always a reliable indicator of protein changes (13). Especially when studies include a temporal aspect, delays between mRNA and protein can combine with other regulatory mechanism to create a more complex relationship (14). Other processes affecting protein dynamics include the inherent stability of proteins, and degradation through the Ubiquitin Proteasome System (UPS), which can also be used by the cell to actively regulate protein availability. The actual function of a protein is additionally not only dependent on protein levels, but can also be affected by post-translational modifications (PTMs), such as phosphorylation or regulatory ubiquitin-chains, or through subcellular localization (15).

Proteome-wide studies of the cell cycle have previously been performed on asynchronous cells by Lamond and colleagues, separating cell cycle phases using centrifugal elutriation (16) or separation of via a combination of DAPI and phospho-histone-H3 on fixated cells (17). Other high-throughput protein studies on the cell cycle utilizing synchronization methods have studied mitotic phosphorylation patterns (18,19) protein stability (20) and protein interactions (21).

1.2 THE UBIQUITIN PROTEASOME SYSTEM

The conjugation of ubiquitin, while primarily used as a degradation signal in the UPS, has been found to signal a broad variety of regulatory functions and has been implicated to play a role in a wide variety of diseases. Its major role, using the classical ubiquitin conjugation machinery, is to direct the activity of the proteasome system and regulate protein degradation. The complexity of different chain linkages, branching of ubiquitin chains, and ubiquitin-like proteins shows that the system is involved in many more types of regulatory control.

The conjugation machinery involved in ubiquitination consists of various subfamilies. The E1 ubiquitin-activating enzymes, the E2 ubiquitin conjugating enzymes and E3 ubiquitin ligases work together to conjugate ubiquitin with targeted proteins, while the deubiquitylating proteins (DUBs) facilitate deconjugation of ubiquitin, both as a regulatory mechanism but also to recycle ubiquitin from the proteasome (22) (**Figure 2**).

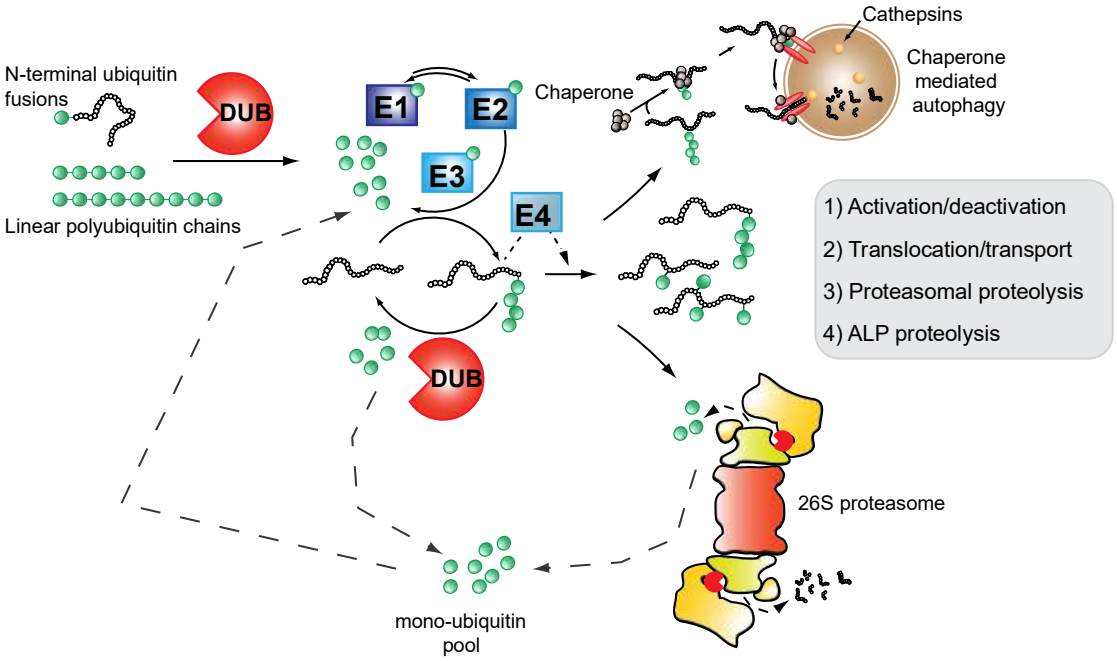

Figure 2. The Ubiquitin Proteasome System

The small protein Ubiquitin was first identified in 1975, and received its name from its ubiquitous expression in all tissues. In 2004, the discovery of ubiquitin-mediated degradation was awarded the Nobel Prize in Chemistry (22). After its initial discovery, a multitude of functions and regulation of processes have been attributed to the system. The multitude of genes dedicated to the UPS system, over 4% of all known genes in the human genome, indicates its extensive use as a regulatory system, not only limited to housekeeping functions.

Four genes produce the ubiquitin protein, UBA52, RPS27A (ribosomal protein-conjugated monoubiquitin), UBC and UBD (linear polyubiquitin). A number of Ubiquitin-Like Proteins (UBLs) have been identified, in size and structure similar to ubiquitin but not responsible for degradation. These UBLs have various roles, for example ISG15 is activated by interferon signaling and conjugates various substrates in immune control (23). The Small Ubiquitin-Like Modifier (SUMO) family of UBLs is comprised of four paralog genes, and involves a number of E1s, E2s and E3s. Sumoylation does not lead directly to degradation, but has been found to play a role in cell cycle control, localization signaling and indirect protein stability regulation through recruiting E3 ligases (24).

The E1 and E2 enzyme families are relatively small, with nine recognized E1 enzymes and around 40 E2 enzymes. The role of the E1 ubiquitin-activating enzymes is to form a covalent thiol-ester bond with ubiquitin, priming it for transfer. The ubiquitin is then transferred to the E2 conjugating enzyme which, in a concerted reaction with the E3 ligase, conjugates it to the substrate. Most substrate specificity is thought to lie in the E3 family, with many E3 enzymes lacking enzymatic activity acting instead as adapter proteins or substrate recognition proteins. Over 600 E3 Ligases have been recognized (22).

The three major groups of E3 families are the RING-(Really Interesting New Gene)-Finger-Proteins (RNFs), the U-Box-containing proteins, and the HECT (Homologous to the E6AP Carboxyl Terminus) Family of proteins. The RING-, HECT-, and U-Box domains are responsible for binding to their corresponding E2 protein, and the other domains have substrate specificity, enzymatic and other adapter functions. While most HECT-family proteins have ligase activity, most RNF proteins only act as adapters or substrate-recognition factors (25).

A common complex form of E3 activity is the Cullin-Ring-Finger (CRL) complex, which contains specialized E3 proteins as adapters, substrate recognition factors and effectors. The most well-known type of CRL complex is the group of Skp1-Cullin1-F-Box (SCF) complexes involved in cell cycle regulation (26).

The E3 Ubiquitin adapter BTBD1 belonging to the CRL3 complex was identified in 2001, together with BTBD2 as a novel family of genes containing the BTB-Domain The BTB-Domain is found to be evolutionary conserved in many species, not only on ubiquitin-related genes. It is thought to facilitate homo-dimerization, and is usually accompanied by a Kelch-domain and/or a zinc-finger-domain, neither of which were found in BTBD1 or BTBD2 (27). A specialized form of CRL complexes have been identified, called CRL3, combining Cullin3, Rbx1, and a BTB-domain-containing protein (26). The BTB-domain is evolutionary conserved

and was first found in Drosophila melanogaster. Known BTB-domain-containing proteins belong to the KBTBD- and KLHL-families (28), but potentially BTBD1 also could be part of the CRL3-complex; affinity-mass-spectrometry experiments suggest a high interaction between BTBD1 and Cullin3 (29).

It has additionally been found that BTBD1 interacts with Topoisomerase-1 (TOP1) (30) and that silencing of BTBD1 induces caspase 3 and caspase 7 activity (31).

An additional family with only two members, the E4 ubiquitin ligases has been found to facilitate multi-ubiquitin assembly and editing (32).

The primary ubiquitination event is usually mono-ubiquitination which enables poly-ubiquitination and the creation of chains. The ubiquitination takes place through a conjugation of the C-terminal peptide of ubiquitin and a lysine peptide on the target substrate. Ubiquitin itself has seven different lysine (K) residues (K6, K11, K27, K29, K33, K48, and K63), and the lysine used in further poly-ubiquitination confers various signals (**Figure 3**). The K48 lysine chain signals for transport to the proteasome and degradation, while the K63 chain linkage is involved in trafficking and cellular signaling processes (22). Other chains have more recently been investigated, and many of their roles are not entirely elucidated (33). K11 chain linkages are known to be utilized as a degradation signal in mitotic exit by the APC complex (34,35) and K27 chain linkages has been implicated in RNF168-mediated ubiquitination of chromatin at DNA damage sites (36).

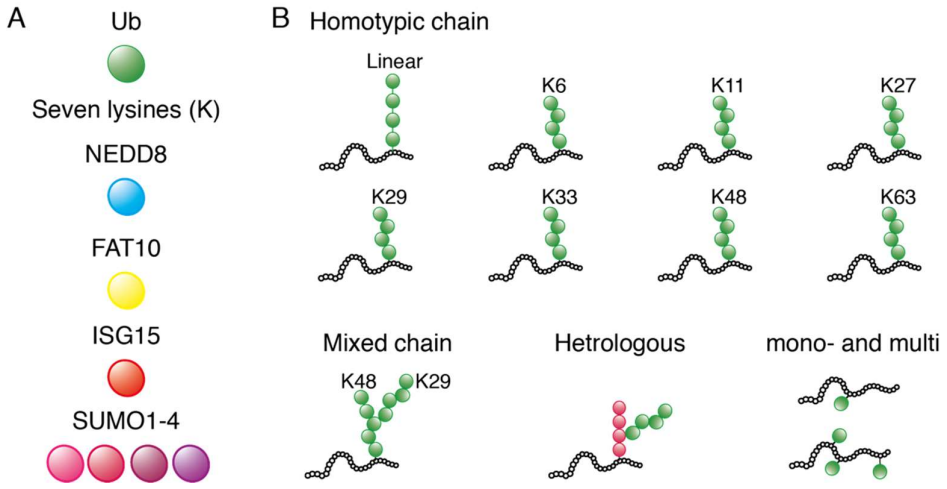

Figure 3. Ubiquitin Chain Linkages

The family of deubiquitylating enzymes is a subfamily of ubiquitin proteins responsible for hydrolyzing the ubiquitin group or ubiquitin-like proteins (UBLs) from target proteins. This family comprises around 100 proteins in five subfamilies, with the largest subfamily being the Ubiquitin-Specific Proteases (USPs) that deconjugates bound ubiquitin (37). The other cysteine

protease families are the Ubiquitin-Carboxyl-terminal Hydrolase (UCH) family, the Machado-Joseph disease protein domain (MJD) proteases, and the Ovarian Tumor protease (OTU) family. In addition, there is a family of metalloproteases, the MPN+/JAMM family, that employs a zinc ion for proteolytic cleavage (38). Other notable subfamilies include the SUMO/Sentrin-specific (SENP) proteases, which deconjugate the SUMO UBLs. Ulp1 and Ulp2 also perform this function, although they were not originally classified as SENPs. An exception is SENP8 which has been found to be responsible for NEDD8 (a Ubl) deconjugation (24).

Deconjugation is used by the cells not only to recycle ubiquitin at the proteasome but also to achieve regulatory control over protein turnover. The balance between the conjugation and the deconjugation machinery can be under fine control, allowing cells to achieve quick changes in protein amounts via ubiquitin system regulation.

Different ubiquitin system pathways have been proposed as therapeutic drug targets in a number of diseases. In many cases of neurodegenerative diseases, such as Parkinson's disease, Alzheimer´s disease and Huntington's disease, part of the disease profile is the accumulation of plaques of proteins. This could be explained by a dysregulation of the UPS system, which makes it a possible target for therapeutic intervention (39).

In cancer biology, the most studied example of the ubiquitin system is the interaction of MDM2, an E3 ligase, and p53 a tumor suppressor. p53 is constantly degraded by the UPS system until signaling changes the balance. Cancer cells need to deregulate this balance, either by functionally inactivating p53 itself or by overexpression of the MDM2 protein (40). This makes this balance an interesting therapeutic target in non-p53-inactivated cancers.

Another therapeutic strategy exploiting the ubiquitin-proteasome system focuses on the class of proteasome inhibitors. Since cancer cells proliferate more frequently and can utilize degradation pathways to inhibit tumor suppressors and apoptotic signaling, proteasome inhibitors have performed well in preclinical and clinical trials. The most successful one so far is a proteasome inhibitor called Bortezomib (Velcade®) (41), targeting the chymotrypsin-like region of the 26S proteasome. Over 200 clinical trials have been performed so far with Bortezomib, and many similar compounds with varying efficacies have been developed (41).

A recent targeting strategy aimed at inhibiting the proteasome is the use of VLX1570 from the Linder Group at Karolinska Institutet. VLX1570 is an analogue of their earlier deubiquitinase inhibitor b-AP15 that claim to target the two proteasome-associated deubiquitinases UCHL5 and USP14 (42). The role of UCHL5 and USP14 is to remove ubiquitin itself from the proteasome in order to recycle it back into circulation. VLX1570 is developed as a more potent USP14-specific inhibitor and is currently in clinical trials (43).

1.2.1 Ubiquitome-Wide Studies

In recent years, a number of large-scale projects have been published describing aspects of the ubiquitin-machinery using different approaches. In 2009, the Wade-Harper lab published a

paper describing a high-throughput immunoprecipitation/mass-spectrometry approach to characterize the targets and associated proteins of the deubiquitylating family of enzymes (44). They have since, together with the Gygi laboratory, developed a proteomic approach to identify ubiquitination sites using mass spectrometry to identify "diGly remnants", and performed these experiments with and without the use of proteasome inhibitors to differ between degradation and regulatory signals (45) or a NEDD8 modification since NEDD8 leaves a diGly remnants on the peptides.

More recently, in 2017, a study by the Vermeaulen group published a paper describing the interaction landscape to ubiquitin chains using uncleavable diubiquitin probes for all the specific chains, and characterizing interactors in three cell types: HeLa cancer cells, embryonic stem cells, and neuronal precursor cells (46).

1.2.2 The Ubiquitin System in DNA Damage Response Pathways

Maintaining DNA integrity is essential for cellular proliferation, and the cell has a multifaceted and highly redundant response machinery to respond to DNA damage (47). The response depends primarily on the specific kind of DNA damage, be it single- or double-strand breaks, bulky adducts or other lesions, but the response can also be affected by the state of the cell.

Common inducers of DNA damage are reactive oxygen species, UV-induced damage from sunlight, genotoxic chemicals and ionizing irradiation. The DNA damage types are multiple; mismatching base pairs, adducts and single-strand or double-strand breaks. As genomic integrity is crucial, especially for dividing cells, the cell employs a variety of strategies to identify and deal with DNA lesions (48). In dividing cells, the DNA damage response (DDR) also influences the cell cycle, by halting the cell at the most appropriate checkpoint (49). The UPS has been implicated to be involved in many DDR pathways and processes, both in a degradation capacity but also through regulatory functions (50–53).

PCNA is a major enzyme in control of DNA replication and DNA repair control and recruitment. It is a trimeric protein encircling the DNA strand and is essential for replication activation. In replication, PCNA is sumoylated by Siz1/UBE2I. After DNA damage occurs, PCNA is activated by mono-ubiquitination by RAD18 to promote translesion synthesis (TLS). To turn off this process, mono-ubiquitylated PCNA is targeted for ISG15-ylation, which then recruits the deubiquitylating enzyme USP10 which removes the mono-ubiquitin. The ISG15 protein is later cleaved off for reuse by USP18 (54). More recently, ISG15 has also been shown to be induced upon various sources of DNA damage (55).

The cell-cycle dependent availability of genetic material can affect the process utilized to repair specific sites, such as the balance between homologous recombination (HR) (56) and non-homologous end-joining (NHEJ) (57) in response to DNA double strand breaks (DSBs) (58–60). NHEJ repairs DSBs through direction ligation, while HR repairs the damaged site through recombination with a homologous segment of DNA, often a sister chromatid. The processes of HR and NHEJ inhibit each other, and the initiation and recruitment of either is part of a complex machinery.

At double-strand breaks, complexes recognizing DNA lesions recruit the ATM Kinase which phosphorylates H2AX. γ-H2AX then recruits various other proteins, among them the E3 Ligase RNF8, which, together with the E2 conjugase UBE2N and RNF168 polyubiquitinates the histone (61–63). This ubiquitination is of the K63-chain variety and has no degradation effect, but recruits 53BP1 and BRCA1, two major players in double-strand break control (63). Another histone modification necessary for 53BP1 recruitment is methylation at H4-K20 which is proposed to be mediated by PCNA (64) but this is still under debate. BRCA1 primarily promotes homologous recombination to repair the double-strand break while 53BP1 promotes non-homologous end-joining. BRCA1 uses the E3 ligase UHRF1 to signal for degradation of RIF1, an essential cofactor to 53BP1. This is performed in S-phase when the cell preferentially wants to repair double-strand breaks using homologous recombination (65). Sumoylation also performs multiple roles during double-strand break repair, including the de-sumoylation of KAP1 to release CDH3 in order to relax the chromatin. BRCA1 has also been found to be regulated positively through direct sumoylation (66). The deubiquitinating enzyme OTUB1 has been identified to inhibit RNF168-mediated poly-ubiquitination at DSB sites, through interacting with UBE2N, the E2 conjugase affiliated with RNF168 (67). Another deubiquitinating enzyme, USP1, is involved in the repair choice at DSBs, and it has been shown both to remove K63-ubqiquitin chains from histones at the control of the APC-complex (60), and through deubiquitinating of PCNA together with UAF1 (68).

1.2.3 The Ubiquitin System in Cell Cycle Control

The control of the turnover of the cyclins as well as CKIs is performed by the ubiquitin-proteasome system. Two major ubiquitin system complexes regulate cell cycle control, the SKP1-CUL1-F-Box-(SCF)-complex and the anaphase-promoting complex/cyclosome (APC/C) (69,70). In the SCF complex, the F-box proteins confer substrate specificity and are variable; currently around 70 have been found. The most important F-box proteins in cell cycle control are SKP2, FBXW7 and β-TRCP (FBXW1) (71,72). SKP2 is a common oncogene in cancer development, as an overexpression of it leads to decreased stability of the CKI and tumor suppressor p27 (CDKN1B). In normal cell cycle progression, SKP2 acts in S-phase and G2-phase to destabilize negative regulators of cell cycle progression (73). FBXW7 on the other hand is a negative regulator of the cell cycle and has been found to be deactivated through mutation in many cancers. β-TRCP is a versatile protein which has been found to regulate stability of both cell cycle drivers and negative regulators, and has been seen to be both overexpressed and mutated in cancers (71).

The APC complex is composed of 13 invariable components, the APC proteins, as well as an effector protein. The effector proteins of the APC-complex are also variable, with two of the most important proteins being CDC20 and CDH1, which act to promote sister chromatid separation and mitotic division. The APC-complex stays upregulated during G1-phase to negatively regulate mitotic cyclins (71). The Fucci System utilized in this thesis utilizes degradation sites recognized by the SCF- and APC-complex respectively in order to act as a reporter of cell cycle progression (6).

1.3 THESIS RATIONALE

The Ubiquitin Proteasome System and the Cell Cycle are two major regulatory systems that have been studied for decades. They are complex in both the sheer number of proteins involved, but also due to interaction between the systems and other cellular processes. This complexity suggests there is still a lot to be discovered and that novel approaches not only in data acquisition methods such as mass spectrometry, RNA sequencing and high content imaging, but also in analysis methods, needs to be employed to unravel the intricacies of these regulatory systems. As high-throughput and high-content methods are getting more advanced and more accessible to the scientific community, I believe the integration of results from modern methodologies will provide additional insights into these basic cellular processes.

2 THESIS AIMS

The cell cycle and the UPS are two highly studied biological systems with well-known regulators. As the advent of modern high-throughput and high-content methodologies have enabled new systems biology approaches to study biological processes, it has also introduced challenges and opportunities in comparative data analysis.

The central aim of this thesis is to study how the cell cycle affects gene expression, protein availability, and protein function, and how the ubiquitin proteasome system is involved in this essential cellular process, while developing novel methods of visualizing and analyzing high-throughput and high-content data.

Specific Project Aims:

Project I: To characterize gene expression profiles throughout the cell cycle, and introduce novel analysis methods to investigate family-wise patterns of oscillation with a focus on transcription factor control.

Project II: To characterize the proteome, the phospho-proteome, and the subcellular proteome over the cell cycle, and investigate the interplay between different cell-cycle regulated systems using novel visualization and analysis methods.

Project III: To characterize the effect of the individual genes in the ubiquitin proteasome system on cell viability, DNA repair pathways and cell cycle control, and the interaction between them, in order to identify novel mechanisms.

Project IV: To develop a highly adaptable vector system for genetic manipulation using shRNA, overexpression and CRISPR/Cas9, a toolbox enabling flexible in-house high-throughput and targeted research.

3 METHODOLOGICAL APPROACHES

3.1 BIOLOGICAL MODELS

3.1.1 Cell Line Models

Throughout the studies, various cell line models were utilized, both unmodified but also integrated with genetic modifications, either to modify the expression of specific genes or to provide reporter system functionalities. For **Study I** and **II**, the cervical cancer cell line HeLa (74) and the osteosarcoma cell line U2OS (75) was utilized, both genetically modified with the Fucci cell cycle reporter system (6). In **Study III**, Fucci U2OS cells (kind gift from Dr. Masai at the Tokyo Metropolitan Institute of Medical Science, Japan) were utilized in siRNA characterization studies, and were also stably integrated with an inducible shRNA expression system targeting a multitude of genes. In **Study IV**, stable shRNA Fucci U2OS cell lines were utilized, as well as multiple glioblastoma multiforme cell lines; U87MG (ATCC), M059K, GBM-H18, U343MG (kind gift from Prof. Monica Nister, Karolinska Institutet), as well as GBM stem cell lines U3013 and U3047 (obtained from HGCC, Uppsala University).

3.1.2 Orthotopic Zebrafish Model

An orthotopic model of glioblastoma multiforme growth in zebrafish hindbrain was utilized in **Study IV** to validate the use of multi-fluorescent viability experiments, using four U87MG GBM cell lines expressing four different fluorescent reporters co-expressed with four unique shRNAs. The orthotopic model was performed as developed by Pudelko et al (76).

3.1.3 Cell Cycle Reporter Systems

The Fucci System consists of two peptides of hCDT1 and hGeminin that are targeted and degraded by the cell-cycle oscillating complexes SCF and APC. The peptides are truncated and limited to the peptide sequence recognized by the APC and SCF-complex. The truncated peptides are conjugated to two fluorescent markers, originally mKO2 and mAG (6), which enables cell cycle classification and separation without any need for fixation or addition of chemicals. The Fucci reporter system integrated into our down-regulatory plasmid system utilizes the green fluorescent protein Clover instead of mAG (77).

3.1.4 Methods of Gene Silencing and Expression

Throughout the present studies, various methods of modifying gene expression were utilized. siRNA libraries were used in order to characterize a large number of gene silencing phenotypes on cell lines in **Study III**, and to further validate some of these findings, stable integrated tetracycline-inducible shRNA constructs were inserted lentivirally into the genetic sequence of various cell lines in **Study III** and **Study IV** (78). CRISPR/Cas9 functionality was additionally integrated in the plasmid system developed in **Study IV**.

3.1.5 Plasmid Development

Through the studies a suite of plasmids was created, both for overexpression of tagged exogenous genes, and to silence genes through shRNA or CRISPR/CAS9. Many different methodologies were utilized for creation and modification of genetic vectors, such as subcloning of PCR-amplified genetic fragments, subcloning from synthesized genetic sequences, or ligation of annealed oligonucleotides. For overexpression purposes, plasmids were separated into pENTR and pDEST plasmids, with Gateway Cloning (79,80) utilized to move sequences into lentiviral-compatible vectors for use in cell experiments. As all silencing vectors and pDEST vectors contained repetitive elements, recombination-deficient Stbl3 *Escherichia coli* cells were used to amplify these vectors. For isolation and amplification of pDEST vectors containing the toxic ccdb-region, ccdb-resistant *E. coli* cells were used. The plasmids were created with a modular design philosophy, to enhance flexibility and ease of use. **Study IV** describes in detail the specific molecular biology procedures utilized for the plasmid systems used throughout the studies comprising this thesis.

The plasmid systems utilized a collection of fluorescent proteins (FPs), distributed along the visual spectra (81). The field of FPs have progressed far since the initial discovery of GFP (82), and directed and random mutagenesis has produced FPs with different spectral properties, many of which are available in our system (6,77,83–88), as well as two different luminescence reporters (89,90). For acquiring comparable data of FPs for visualization and comparison, the FPBase online database proved useful (91).

3.2 DATA ACQUISITION

3.2.1 Cell Biology Techniques and Phenotypical Characterizations

This section summarizes essential assays and experimental techniques utilized throughout the thesis in characterization of cellular processes, both observationally without intervention, but also after external regulation of genes and proteins or chemical intervention.

3.2.1.1 Viability Assays

Throughout the studies, various methods of measuring cellular viability and proliferation were employed, from plate-measurements, colony-formation assays and counting individual cells. Resazurin viability assays measures cellular viability in an entire well through measuring the color change of Resazurin in response to aerobic respiration (92). A Trypan Blue colony-formation assay was used to verify effects on cellular viability in inducible shRNA cell lines in **Study III**. The highest resolution of cellular viability, the exact count of cells was achieved via two methods, through flow cytometry in **Study IV** and through automated counting of segmented DAPI-stained nuclei in microscopic images in **Study III**.

3.2.1.2 Cell Cycle Characterization Methodologies

To characterize the cell cycle distribution in a population of cells, two distinct methods were utilized through **Study III** and **IV**. Categorization using the Fucci system was performed by

quantifying the expression of mKO2 and mAG fluorescent proteins, and categorizing each cell as either single-positive for one (G1- and G2/M-phase respectively) or double-positive (S-phase). This was performed either by using flow cytometry and through automated classification of high-throughput microscopic images (**Figure 4**).

As the Fucci system is a reporter of the cell's regulatory cell cycle progress, the state of the DNA is a reporter of the mitotic cell cycle. We utilized a combination of DAPI quantification and quantification of replication-integrated EdU on microscopic images to accurately categorize cells on events regarding DNA content and replication status. When DAPI staining was utilized for chromatin quantification, it was always preceded by an hour-long incubation with RNAse I and PBS-washes, in order to remove possibilities of intercalation of DAPI to RNA. EdU is a thymidine analog which contains a free alkyne group, allowing for click-chemistry-based conjugations of azide-conjugated fluorophores post-fixation (93–95).

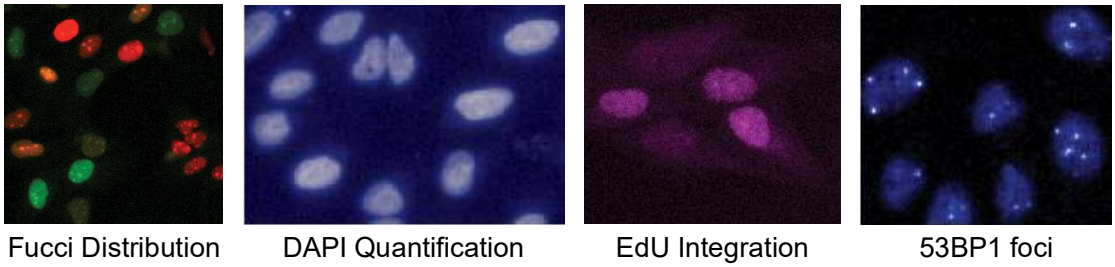

Figure 4. Example fluorescence images of Fucci, DAPI, EdU and 53BP1 foci

3.2.1.3 DNA Integrity, Damage and Response Assessments

In order to characterize cellular phenotypes related to DNA integrity and the DDR, multiple methods were utilized.

To investigate DNA replication speeds, a fiber assay was utilized in **Study III**, wherein cells were incubated with two sequential nucleotide analogs (CIdU and IdU), and single molecule DNA fiber lengths were measured using fluorescence-conjugated secondary antibodies. The distribution of fiber lengths corresponds to the distribution of replication capacity with the nucleus.

To investigate DNA integrity in **Study IV**, a Single Cell Gel Electrophoresis (COMET) assay was used, wherein low melting agarose encapsulates the cells followed by electrophoresis. Fragmented DNA then travels further and creates the titular comet tails which lengths are measured and compared to the nuclei.

Antibodies against DDR proteins were also utilized, both using western blotting but also using immunofluorescence techniques. Using immunofluorescence, foci formation patterns of the DSB-recruited enzymes BRCA1 and 53BP1 could be counted as an indicator of the cellular

response to DSBs, and the choice between homologous recombination and non-homologous end-joining. Integrated intensities of the two proteins was also measured to control for availability.

3.2.1.4 Interaction studies

Two methods were used to investigate protein-protein or protein-DNA interaction. Proximity Ligation Assay (PLA) was used to verify the interaction between the E3 Ubiquitin Adapter BTBD1 and TOP1 in **Study III**. PLA works by conjugation of two complementary oligonucleotides to two secondary antibodies, and an in-cell amplification and binding of a fluorescent reporter when these two oligonucleotides are co-localized. Another method of protein interaction, chromatin immunoprecipitation (ChIP), was used to characterize the BTBD1-silenced-mediated binding of TOP1 to transcription start sites of the genome.

3.2.1.5 High Content Imaging-Specific Methodologies

For performing high-content imaging experiments in **Study III**, various liquid distribution and plate handling equipment were utilized. For initial siRNA library generation, siRNAs targeting UPS genes were collected from a genome-wide library using an Echo Liquid Handler. For plating of cells, reagents, and washes in 384-well plates, an 8-tube peristaltic MultiDrop dispenser was utilized. Images were acquired on an Operetta HighContent Imaging System (PerkinElmer) and an ImageXpress HighContent Microscope (Molecular Devices), connected to robotic plate handlers (Thermo). Image analyses was performed through batch processing at the UPPMAX data cluster, and output data stored in a SQL database. Data analysis was performed in the software R (96) using the *dplyr* package (97) and figures were prepared with the *ggplot2* package (98).

3.2.2 Transcriptomics, Proteomics, and Phospho-Proteomics

Transcriptomics and proteomics approaches were used primarily in **Study I** and **II** on Fucci cells sorted by flow cytometry in order to objectively study the cell cycle oscillations of mRNA and proteins. In **Study I** we utilized transcriptomic analysis to detect mRNA expression patterns over the cell cycle in two cell lines, U2OS and HeLa. In **Study II**, cell-cycle oscillating proteins were investigated, and phospho-proteomics were used to additionally characterize cell-cycle dependent phosphorylation events. In **Study III**, mass spectrometry was utilized to investigate proteomic changes after shRNA-mediated BTBD1 silencing. Cellular material was labeled using an isobaric mass tag TMT10plex method (99), and a TiO2 phosphopeptide enrichment method was used prior to phospho-proteomic mass spectrometry. Proteomic Analysis was performed with the softwares MaxQuant (100), Andromeda (101), Perseus (102), and R (96). Raw data from transcriptomic experiments was submitted to the NCBI GEO repository (103) and raw proteomic data was deposited at the PRIDE repository (104).

3.3 DATA ANALYSIS

3.3.1 High Content Imaging Data Analysis

For all analysis of immunofluorescent images, the software CellProfiler was used (105,106). CellProfiler contains a library of image analysis algorithms, from image organization, image normalization, segmentation and feature extractions. After feature extraction, categorization of cells and summarization of results was performed in R, starting with the per-object output files from CellProfiler. The process from acquired images to analysis of data is summarized in **Figure 5**.

Choice of features to utilize for each fluorescent marker was chosen according to the mechanism and function of each marker or reporter. For DAPI quantification, where the aim was to quantify the cellular content of DNA, integrated intensities were used.

Normalization was performed either before feature extraction to remove variable background noise arising from optical effects and uneven backgrounds, or in R on the extracted feature descriptors, where negative control wells spread throughout the plates were used to correct for batch-wise, inter-plate, or intra-plate variations.

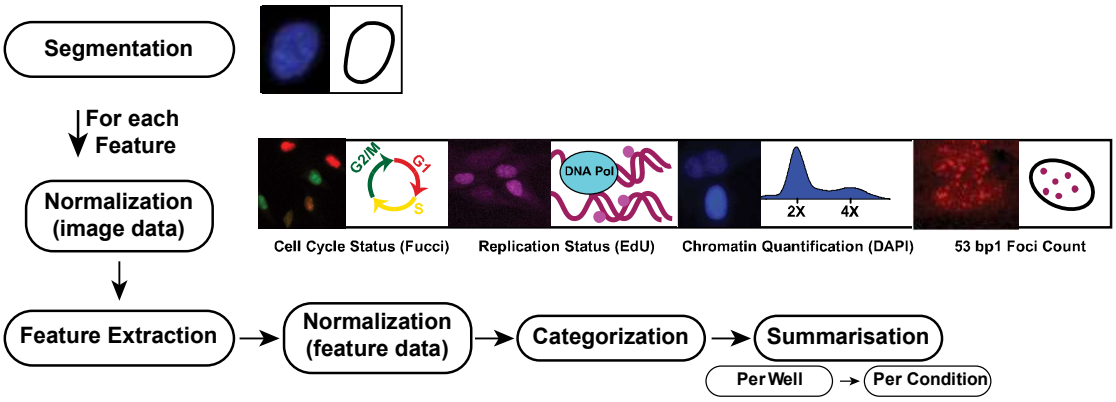

Figure 5. High Content Image Analysis Pipeline

3.3.2 Comparison of Tri-Group Relative Data Utilizing a Polar Coordinate System

To visualize and compare cell cycle oscillation patterns an algorithm was developed, called TriComp. It utilizes a polar coordinate system to plot the relationship between a quantitative variable measured in three groups, while ignoring the average value. This enables a structured approach of plotting the thousands of data-points generated from transcriptomic or proteomic experiments. The radius corresponds to the logarithmic fold change difference between the three groups, and the degree angle corresponds to the specific relationship between the three groups. The triangularization equations and pedagogical examples of how the algorithm functions is summarized in Suppl Figure 1 of **Study II**.

3.3.3 Comparison of Tri-Group Distribution Data Utilizing a Triangular Coordinate System

To visualize and compare cell cycle distribution patterns in **Study III**, another algorithm was developed. The algorithm produces coordinate values in a triangular coordinate system, where each point corresponds to a distribution of three groups, summing up to 100%. An imagined line perpendicular to each side of the triangle directly corresponds to the percentage of each of the three cell cycle phases.

3.3.4 Statistical Analyses

In **Study I**, in order to calculate statistically significant oscillations of transcripts over the cell cycle, a Generalized Linear Model using a negative binomial model was used, from the R package *edgeR* (107). Similarly, in **Study II**, to calculate p-values of statistical significance on proteomic and phospho-proteomic results, Analysis of Variance (ANOVA) was performed for each unique protein or phosphopeptide.

Additionally, throughout the studies, different enrichment analyses were performed. When comparing a distribution of factors to an encompassing parental distribution, log-likelihood ratios were computed using the XNomial package in R (108). When comparing separate distributions, fisher tests were performed in R, with higher values using a Monte Carlo simulation.

Gene Ontology enrichment analyses was performed in **Study I and II**. When comparing a set of genes or proteins to Gene Ontology gene sets (109,110), the online resource DAVID Functional Annotation Tool was utilized (111,112). When comparing gene lists to individual gene sets, such as the cell-cycle-specific GO-terms in **Study I**, this was performed in R or Microsoft Excel. In Study II for identifying enriched kinase motifs, the motif prediction tool in Perseus (102) was used.

When comparing normally distributed quantitative values of two groups, Student's T-tests were utilized.

4 RESULTS

4.1 CHARACTERIZING CELL CYCLE OSCILLATIONS OF MRNA, PROTEINS, AND PROTEIN FUNCTION.

In **Study I**, we characterized the cell cycle oscillating transcriptome in unsynchronized cells. Aside from confirming major cell cycle regulatory pathways to be transcriptionally regulated, both through known examples and through GO-term enrichment analysis, we could also show the inherent link between the circadian rhythm and the cell cycle.

In order to visualize and categorize cell-cycle oscillating transcripts, we developed a novel algorithm using a polar coordinate system: TriComp. For a TriComp graph, the radius (r) corresponds to the intensity of the cell cycle oscillation, and the angle (θ) corresponds to the type of relationship between the three groups. In the TriComp comparison graphs, a subsequent pattern does not necessarily mean a later expression peak, as the results focus on the relationship between the three groups, and not which individual phase has the maximum expression. A basic subdivision of the relationships was used to divide all hits into G1-, G1+S-, S-, S+G2-, G2-, or G1+G2-upregulated.

Overall, the majority of cell-cycle oscillating transcripts were either moderate level G1+S-specific or high level G2-specific. Stratifying the data on transcription factors revealed an overall similar patterns for most TF families. An interesting exception was the E2F family of transcription factors, which were all upregulated in S-phase, either alone or with a corresponding upregulation in G1-phase or G2-phase. We investigated known targets of the E2F1 transcription factor, and saw a highly significant corresponding pattern of E2F1-targets, compartmentalized between G1+S- and S+G2-expresstion patterns. Comparing U2OS and HeLa cells show a conserved pattern of cell cycle transcriptional regulation, with very few hits having different cell-cycle enrichment phases, but an investigation of U2OS- and HeLa-specific oscillating transcription factors and a GO-term enrichment analysis revealed these to be enriched for different GO-terms involved in differentiation and developmental programs.

In **Study II** we used the biological model and the TriComp algorithm from **Study I** to characterize the cell-cycle dependent behavior of the proteome and phospho-proteome over the cell cycle, as well as investigate cell-cycle dependent changes in subcellular localization. Results were also compared to published data on gene essentiality acquired in a comparison between ten CRISPR/Cas9 survival screens (113). The TriComp algorithm was used to compare between different modes of data, and to characterize similarities and differences between transcriptome, proteome, and external databases.

A total of 3355 cell-cycle oscillating proteins were identified, with the majority exhibiting an upregulation in either G1+S or G2, with the highest fold-change differences belonging to G2-specific proteins. Enrichment analysis for GO-terms revealed a high enrichment for cell cycle processes, with many known cell-cycle regulated proteins appearing high up in the hit-list. The

same analysis on each TriComp group also revealed functional differences for each specific group.

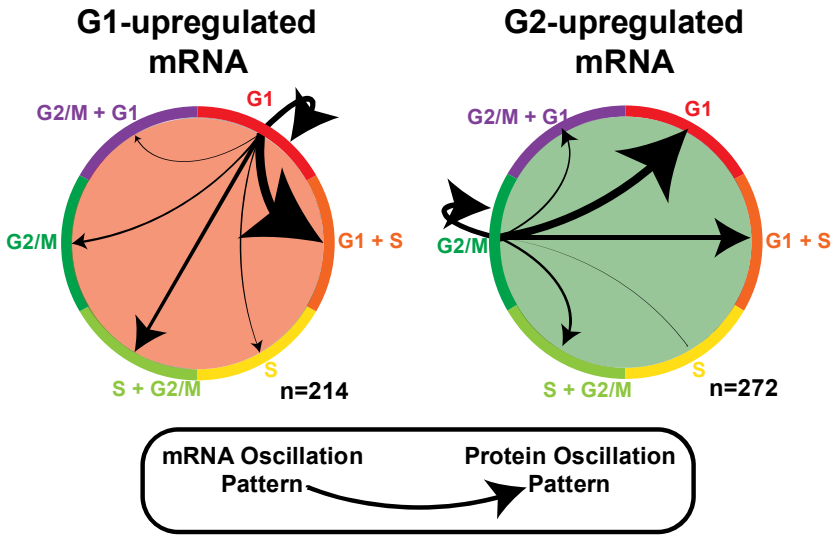

Figure 6. Examples of comparing transcriptional and proteomic regulation over the cell cycle

When comparing transcriptomics and proteomics over the cell cycle we identified a high degree of correlation, either with transcripts and proteins exhibiting the exact same cell cycle expression pattern, or with proteins exhibiting an immediate subsequent pattern. Stratifying for each specific cell cycle phase revealed differences between each group of mRNA expression patterns. For example, G1-upregulated mRNAs mostly corresponded to G1- or G1+S-upregulated proteins, but G2 upregulated mRNAs showed a tendency to correspond to proteins upregulated in the following G1 phase (**Figure 6**).

Phosphatases and kinases were found to have statistically different oscillation profiles and phospho-events were characterized over the cell cycle through the use of TiO_2 enrichment followed by mass spectrometry. 3317 phosphopeptides were found to oscillate over the cell cycle, and kinase motif recognition was utilized to describe motifs enriched for certain TriComp patterns.

The proteomic analysis was also repeated but following fractionation of cellular material into soluble and insoluble fractions. By filtering out proteins which did not oscillate in the general proteomic results, but exhibited diametrically opposite patterns on the soluble and insoluble cell cycle, a list of 169 proteins was identified to change solubility state over the cell cycle. A clustered group of these was the MCM family, which was upregulated as insoluble in S-phase and as soluble in G1 and G2. Another protein was the S-Adenosyl-methionine donor MAT2A, which was upregulated in soluble G1 phase and in insoluble S and G2 phases. This was validated using immunofluorescence experiments to be due to a translocation from the cytosol to the nuclei following the G1/S checkpoint.

4.2 CHARACTERIZING DYSREGULATION OF THE CELL CYCLE THROUGH SIRNA-MEDIATED SILENCING OF THE UPS

In **Study III** we developed a high-content imaging experiment using a siRNA library against 786 genes of the UPS system, in order to characterize cellular phenotypes regarding cellular viability, the mitotic and regulatory cell cycle, and responses to intrinsic and extrinsic DNA damage. The UPS system has many confirmed and putative members, and we combined information about UPS genes from various databases (37,114–116) prior to designing the experiment.

Four reporter systems were combined in the experiment; Fucci as a measurement of the regulatory cell cycle, DAPI quantification and EdU integration as an indicator of the mitotic cell cycle, and 53bp1-foci counting as a marker of DNA damage response to double-strand breaks. The study was performed at two time points and in parallel with and without external irradiation two hours prior to the first time point. The results showed a broad distribution of phenotypes among the UPS genes silenced, with an enrichment of aberrant phenotypes in the UPS subfamilies comprising a smaller number of genes, both on viability and on the cell cycle. A triangularization algorithm was used to illustrate phenotypical cell cycle results, on a three-group categorization of either Fucci or combined data from DAPI and EdU integration quantification. The study was internally repeated on 20% of the initial study, and the same experimental setup was used on a different siRNA library comprising 78 targets, primarily E2 family members, and on lentivirally transfected inducible shRNA cell lines, primarily targeting DUB family members.

In the primary study, cell cycle phenotypes without external irradiation was primarily directed towards G1- or S-phase enriched phenotypes, with very few siRNA pools exhibiting a G2 arrest. Comparing results from the two time points showed that most affected phenotypes got progressively worse over time, with many moving towards S-phase arrests. Almost every single cell cycle phenotype was accompanied by a corresponding effect on cellular viability.

Through analyzing effects of external irradiation, it was discovered that the percentage of G2 cell cycle arrest phenotypes was significantly higher. Arrow graphs comparing cell cycle phenotypes with and without irradiation revealed many of the siRNA pools causing G2 arrests to either not exhibit any cell cycle phenotype without irradiation, or to exhibit G1- or S-phase enrichments.

The response to DSBs was investigated in the same experiment via 53bp1-foci formation counting, and four categories of phenotypes were detailed using data from all experimental conditions; spontaneous foci formation (early or late), a diminished response to DNA damage, and a normal response but unresolved foci at the latest time point. These four categories were investigated regarding cell cycle effects at all experimental conditions, and two major significant patterns emerged. First, spontaneous foci formation, especially at the early time-point, was often accompanied with cell cycle enrichments towards G1/S-phase. Secondly, a lack of recruitment of 53bp1was often accompanied by a shift in cell cycle arrest patterns, from

no phenotype or G1/S-enriched phenotypes without irradiation towards G2-phase arrests when irradiated.

4.3 BTBD1 SILENCING AFFECTS CELL CYCLE AND DNA DAMAGE REPAIR PATHWAYS, AND IS ESSENTIAL FOR TOP1 FUNCTION

The E3 adapter BTBD1 was discovered in the initial experiment of **Study III** to be involved in viability, cell cycle and DDR response. Without irradiation, silencing of BTBD1 caused a G1/S-phase enrichment of cells, and combined with irradiation this was converted to a G2-enrichment. This was additionally confirmed using an inducible shRNA cell line targeting BTBD1. Additionally, recruitment of 53bp1 upon DNA damage was severely reduced upon silencing of BTBD1, both using siRNA and shRNA, and the ability to perform DNA replication was impaired as measured using DNA fiber experiments.

A phosphoproteomic experiment on BTBD1-silenced cells revealed a profile similar to that of external DNA damage, except for a decrease in phosphorylation of 53bp1. Investigating synergistic effects of BTBD1 silencing with a small chemical library revealed sensitivity to many cell cycle inhibitors but also insensitivity to chemicals related to DNA, such as the compound Camptothecin (CPT), an inhibitor of TOP1, a protein releasing DNA torsional stress during replication and transcription, that BTBD1 has previously been found to interact with (117). A chromatin-immunoprecipitation study of shBTBD1 cells revealed global arrests of TOP1 on transcription start sites, results that were removed upon the co-incubation with CPT. TOP1 was found to be reduced in western blot lysates similarly and synergistically with CPT and shBTBD1, independent of ubiquitin-dependent degradation.

4.4 DEVELOPMENT OF A MULTI-FUNCTIONAL MULTI-FLUORESCENT PLASMID SYSTEM FOR GENETIC ENGINEERING

Study IV describes the design and development of a multi-purpose genetic engineering system and show proof of concept results for the various integrated features. We also validate and characterize two ubiquitin E3 adaptors, ARIH1 and ARIH2, as important for Glioblastoma Multiforme development.

The vector system introduced is comprised of four different kinds of vectors; silencing vectors, pENTR preparatory cloning vectors, constitutive pDEST expression vectors, and inducible pDEST expression vectors. The gRNA/shRNA-compatible gene silencing vector is ready-to-use and lentivirally compatible. The overexpression system utilizes Gateway Cloning, where expression products can be modified in a small pENTR vector flanked by two attL-sites, and later recombined using LR Clonase between two attR-sites in a pDEST vector, which is lentivirally compatible and expression-ready. This allows for a flexible cloning approach, as a gene product can be cloned in the pENTR vector, subcloned to be tagged with different reporter proteins, and then Gateway cloned into different destination vectors for different purposes. There is a large amount of available destination vectors compatible with our system for various purposes, such as recombinant protein production in different species, overexpression for co-immunoprecipitation studies, or for mapping localization within the cell (118).

Our vector system has a number of reporters, protein tags, selection markers, and other features, and a central part of the cloning design is ease of integration of the same reporter into the various systems. For example, a fluorescent protein can be used as a reporter of successful gRNA integration, a N- or C-terminal tag in an pENTR vector, or as a reporter in a destination vector. As similar restriction digestion sites are utilized in the different vectors, the integration of a new feature is easy and efficient.

Fourteen different fluorescent markers are already integrated in the system spanning a wide range of the visual spectra. This allows for a high degree of flexibility in combining fluorescent markers depending on experimental direction or limitations in microscopes available. Having a large array of colors to choose from enables a high flexibility in choosing fluorescent markers for flow cytometry and immunofluorescence experiments.

The system was utilized to investigate the role of two E3 ubiquitin ligases, ARIH1 and ARIH2, on cell cycle and viability. Using multi-fluorescent experiments, they were identified to be essential for cellular proliferation in various GBM cell lines and stem cells.

5 DISCUSSION

5.1 THESIS FINDINGS

5.1.1 The Cell Cycle's Effect on Transcription, Protein Availability, and Protein Function.

In **Study I,** (119), the comparison of transcriptional cell cycle regulation between two commonly used cell lines revealed a high degree of correlation, not only to each other, but to previous studies and to previously identified cell cycle-regulated processes. This emphasizes the importance of cell cycle integrity, and validates the use of cell lines as biological models of the cell cycle. While U2OS and HeLa cells do not actively differentiate into specific post-mitotic cells and are quite homogenous, it was interesting to see that different differentiation and developmental programs could still be seen using after gene ontology enrichment analysis for cell-type-unique transcription factors. The specific GO-terms identified for each cell type related to differentiation and developmental are most likely of lesser importance as many transcription factors can be involved in multiple cellular programs. It is natural that the cells' different heritage could still be shown in the cell-cycle oscillating transcriptional control of these cells.

Other transcriptional networks were found to be oscillating in synchrony with the cell cycle, many with previously identified cell cycle-dependent functions, such as transcription factors like E2F1 (120–122), PAX6 (123–125), Notch signaling (126), WNT/Frizzled (127,128) and the circadian clock (129–132). The circadian clock is additionally known to regulate some of these factors (124,133,134) and has previously been shown to by synchronized with the cell cycle in *in vitro* cell lines (129,130). As these processes was found to be synchronized to the cell cycle in cell line models, this opens up the possibility of other cell-cycle processes being controlled by either the cell cycle directly or by these systems. A limitation of observational studies is drawing conclusions about cause and effect, and unraveling regulation patterns of closely synchronized systems requires targeted research. TriComp proved useful when analyzing distributions of known E2F targets to define a temporal compartment of E2F function, as the relationship between a transcription factor and its targets can introduce a pseudo-temporal method of analysis.

Comparably to the transcriptomic analysis, a proteomic analysis of the cell cycle in **Study II,** (135)**,** confirmed many known cell cycle regulated proteins and corresponded well with GO-terms regarding cell cycle, while providing thousands of significant cell-cycle regulated proteins. Some novel proteins not GO-annotated such as AHRGAP11A, KIAA0101 and TK, had previously been implicated as being involved in cell cycle functions (136), or being degraded in a cell-cycle dependent manner (137,138), but many proteins identified had not previously been implicated in cell cycle biology. Interestingly, GO-terms associated with the cell cycle was more prevalent in S+G2 or G2-upregulated patterns, while G1- and G1+S-upregulated patterns were mostly comprised of previously non-GO-annotated proteins. This

could possibly be due to the G1 phase being a preparatory gap phase. As a majority of cell cycle physical changes happening in S-phase and mitosis, an enrichment in GO-cell-cycle-terms for proteins interacting with these later phases is natural. Still, it is interesting that many proteins do exhibit an upregulation in G1-phase, and enrichment analysis revealed many G1-upregulated proteins to be involved in ribosome biogenesis, a preparatory function for building up cellular material.

Gene ontology enrichment analysis on each different TriComp pattern was useful in defining temporal compartments for different processes. Many processes were unsurprising, such as the fact that "DNA replication"-annotated proteins were upregulated in S-phase (although proteins even carried over into G2-phase). Interestingly, the two different groups of G2-upregulated and G2+G1-upregulated proteins revealed a difference in GO-terms, where "Mitotic Nuclear Division" preceded "Chromosome Separation". Proteins involved in chromatin separation would take longer to be removed than proteins involved earlier in the mitotic process, thus explaining their presence in the following G1-phase.

Comparing mRNA and protein oscillations revealed mRNA fluctuations translated to either a similar pattern on protein levels or one TriComp group forward. A TriComp analysis of relationships (instead of estimating the time-point of maximum expression) was useful here in order to separate proteins upregulated through a single phase from proteins upregulated and remaining elevated into the next phase. Aside from transcriptional regulation which is reflected in the mRNA oscillation patterns, the turnover time of the proteins through stability and degradation affects the protein levels in a specific cell cycle phase and the combined knowledge of mRNA and protein oscillation patterns can hopefully be useful as a resource. The results also highlight the complexity of dynamics between transcriptional regulation, protein expression and stabilization of protein levels in the cell (14).

The conversion of transcriptional regulation to protein level changes is not instantaneous and results regarding G2-specific mRNAs corresponding to G1-upregulated proteins might highlight a preparatory function of the cell prior to mitosis to prepare for functional needs of the daughter cells. During the phases of mitosis when chromatin is condensed transcription is not possible (139), delaying the production of G1-specific proteins further unless prepared for prior to mitosis.

Phospho-proteomic and translocation results from **Study II** in combination with earlier results strengthens the resource aspect of the work, and the translocation-finding algorithm of identifying diametrically opposite cell-cycle oscillation patterns in soluble and insoluble oscillating proteins proved a useful method of identifying proteins changing subcellular localization in response to cell cycle events. Proteins exhibiting differences in soluble and insoluble cell cycle expression could be explained either as a solubility change of larger cellular structures, a solubility change of a protein and its interaction partners, or a translocation between two cellular compartments. Examples of all three reasons were identified, as traces of mitochondrial fission events (140–142) could be seen in the soluble fraction in S-phase, the stabilization of MCM proteins on DNA to enable replication during S-phase (143), and the

translocation of MAT2A from cytosol to nucleus to act as a methyl donor for the cells' increased methyl needs in S and G2 phases (144–146).

Most hit lists from **Study II** were enriched for essential genes for cellular proliferation, based on comparison with CRISPR/Cas9 survival screening studies (113). This enrichment highlights the link between essentiality of function and expenditure of regulation, as a gene coding for a protein with an integral function is more likely to be regulated. The two most integral functions for a cell is unhindered progression through DNA replication and mitosis, and maintaining integrity of the DNA, and these processes has changing needs throughout the cell cycle.

5.1.2 Characterizing Effects of the Ubiquitin Proteasome System on Cellular Viability, the Cell Cycle and the DNA Damage Response

The characterization of the gene-per-gene silencing of the ubiquitin proteasome system in **Study III** revealed a wide and complex array of phenotypes, highlighting the large amount of protein and cellular processes in which the ubiquitin system is present, both in a regulatory capacity and through degradation of unwanted proteins. Differences in subfamily phenotype distributions were identified; E1 enzymes, E2 conjugating enzymes, and deubiquitinating enzymes were enriched for phenotypes compared to E3 ligases and E3 adaptors, both regarding viability and cell cycle. The lower prevalence in phenotypes in E3 family members can be explained by them having many specific roles, many which utilizes the same E1 and E2 enzymes. It is likely that many of these E3-mediated processes are not required in the U2OS cell line or being used for higher-order functions such as cellular organization or tissue- and organ-specific functions. This would explain the lower prevalence of viability phenotypes in this ubiquitin subfamily.

The combined phenotypical results from different reporter systems, from different time-points, and with external irradiation, were utilized together to elucidate effects of each specific siRNA-based silencing of a UPS-related gene. The combination of regulatory cell cycle data with mitotic cell cycle data adds more information when describing a dysregulation into the cell cycle. Using a triangularization method of visualizing the results on cell cycle, the effects could also be compared over time and after irradiation. This revealed that most cell cycle phenotypes enriched for a specific phase are intensified from the first to the second time point, also correlating with viability.

Silencing of UPS genes exhibited various cellular phenotypes, primarily enriched towards G1-phase or S-phase. Upon external irradiation however, many G2-enriched phenotypes appeared, many from proteins previously not having a phenotype. The difference in these phenotypes could be due to synthetic lethality, where two events which the cell is capable of handling together becomes unmanageable. The UPS proteins exhibiting these phenotypes could be involved in a process where the cell has redundancy, or is able to deal with the consequences without stopping cellular proliferation. For a protein involved in DDR or maintenance of DNA integrity, the addition of a secondary problem, such as external irradiation, could overwhelm the repair machinery and lead to an inability to progress through mitosis. The cell has many

processes involved in handling unexpected problems, especially DDR, and an external trigger of problems might be required to reveal an inability to respond. **Study III** also revealed that when silencing UPS genes resulted in a decrease in 53BP1 recruitment to DSB sites upon irradiation, the cell cycle phenotype the following day was more often a G2-phase arrest. This supports the argument that removing proteins essential for DDR could synergize with irradiation to create a cell cycle arrest.

The combination of reporters for viability, cell cycle and DDR response in the same high-content experiment allowed for summarizing and comparing results both regarding specific genes, but also allowed comparing how dysregulation of cellular processes correlate and interact to each other, using the UPS genes as a snapshot of the human genome.

5.1.3 The Role of BTBD1 in TOP1 Function, Cell Cycle Progression, and DSB Repair

In **Study III**, siRNA-based silencing of the E3 adapter BTBD1 exhibited a plethora of phenotypes. It changed cell-cycle enrichment phenotypes upon irradiation, and had one of the highest effects on reducing 53BP1 recruitment to DSB sites. This was additionally confirmed with shRNA-based studies and the HR/NHEJ imbalance was studied further with immunofluorescence and phospho-proteomic experiments. BTBD1 was previously known to interact with TOP1 (30,117), confirmed through a proximity ligation assay and was also found to arrest TOP1 at transcriptional start sites and reduce available TOP1 in western blot lysates similarly to CPT. This points to BTBD1 having an essential role for TOP1 function, although not through ubiquitin-mediated degradation as TOP1 reduction persisted even with the inclusion of MG132.

CPT is used clinically in cancer therapy, and stabilizes TOP1 covalently bound to DNA during the time the DNA unwinds (147). While CPT mimicked the effects of BTBD1 silencing, it did not cause the same pattern of TOP1 arrest at TSS identified through ChIP. This might be due to differences in dynamics. While CPT has a more extreme effect on replication, it is not itself covalently bound to TOP1. As TOP1 is involved to release torsional stress in DNA during replication, transcription, and possibly DNA repair (148–151), the effect of BTBD1 silencing on TOP1 could help explain the various phenotypes on the cell cycle and DSB repair choice observed upon BTBD1 silencing.

5.2 METHODOLOGICAL CONSIDERATIONS

5.2.1 Advantages and Disadvantages of Cell Line Models

This thesis relies primarily on results from *in vitro* immortalized cell line culturing. Laboratory cell lines are advantageous in that their growth is consistent and experimental results are therefore highly reproducible. As the cells belong to a homogenous population, other basic cellular systems are also consistent and reproducible, such as responses to DNA damage, replication, and transcription regulation. Other advantages to cell line studies is the ease and flexibility of use compared to tissue or animal models. On a molecular and intra-cellular level,

cell line research is thus a highly advantageous flexible model system to study protein and cellular biology.

The trade-off is limitations in translation to higher-level functions, such as tissue-level spatial organizations and signaling, and organ- and organism-level functionalities. A caveat that is always present in high-throughput or omics- cell line studies is that an unknown percentage of studied genes and proteins could have functions of higher orders that are not reflected in the model, or have a cell-type specific function in a cell type other than the one studied. In **Study I**, unique HeLa and U2OS cell-cycle regulated transcripts were enriched for developmental GO-Terms, reflecting their difference in tissue origin. Often, transcripts or proteins without a cell-type specific function might not be detected due to low expression, but they could still be detected and even found significantly oscillating due to co-regulation, intended or unintended. Although genomic analysis can be used to study co-localized genes in the same chromosomal vicinity and separate primary expression results to passenger results, spatial co-localization from different genomic loci can also play a role in gene regulation (152). In **Study III**, the lower prevalence of cellular phenotypes upon silencing E3 ubiquitin ligases and adapters can also be explained by many of these having possible functions not necessarily needed in a biological cell line model. As different E3 proteins utilize the same E1 and E2 proteins, this explains the latter's higher prevalence of phenotypes when removing these multi-functional family members. While not in the scope of this thesis, integration of analysis tools on genomic localization and spatial chromosomal localization could prove advantageous to further elucidate the interplay between genes and to identify co-regulated networks.

5.2.2 Studying the Cell Cycle – A Comparison of Methodologies

Studying the cell cycle can either be performed through synchronization of cells and performing an experiment in the time-frame before the cells desynchronize, or with exogenous reporter systems. In this thesis, two methods were utilized to study the cell cycle in asynchronous cells; the regulatory Fucci reporter system and a combination of the chemicals DAPI and EdU to describe the mitotic cell cycle by measuring chromatin content and chromatin replication. In **Study I**, cell-cycle transcriptional oscillation patterns from asynchronous Fucci cells were compared to published data from synchronized cells (10), with a high degree of correlation, although the information procured differed as the TriComp algorithm provided a measure of relationship between three groups.

Fucci as a reporter system provides information on the SCF/APC-complex balance in the regulation of the cell cycle, while DAPI and EdU reports on the physical stage of DNA replication. During the unperturbed cell cycle the reporter systems correlate highly, but when disturbing the cell cycle, the systems can differ. This can be advantageous in elucidating the state of the cell, and to infer possible reasons for cell cycle arrest. Fucci and DAPI/EdU can thus complement each other and the combined information can give a detailed description of cell cycle state.

A limitation of using Fucci or DAPI/EdU methodologies is the need for categorization of stages. In a synchronization experiments, cells can be followed up at certain time-points, a quantitative variable, and the temporal resolution will depend on the methodology and specific biological question. When studying asynchronous cells using either Fucci or DAPI/EdU data, a categorization of cell cycle phases is required, and this requires different strategies in downstream analysis.

5.2.3 A Comparison of Gene Silencing Methodologies

Throughout the studies, different methods of gene silencing were utilized; siRNA libraries, inducible shRNA vector constructs, and single-site targeted CRISPR/Cas9. The choice of gene silencing system depends on the specific use case, and weighing the downside of immediate stress to the target organism versus ease of use, consistency and efficacy.

siRNA libraries are easy to use and scale up to higher throughput. The major downside to siRNA-mediated silencing is its time-dependency, and possible variance between experiments. Stably transduced inducible shRNA constructs on the other hand, separates the immediate stress of viral delivery and genetic insertion from the actual experiment. Successfully integrated cells can be selected using antibiotic selection or through flow cytometry with a fluorescent reporter. After cells have recovered from the lentiviral-mediated integration, cells can be frozen down in homogenous vials, and experiments can be performed again and again with minimal variance arising from the integration method itself. Additionally, due to the shRNA being induced by doxycycline, the same cell line can be used as a negative control, without induction of shRNA, further limiting technical variance. The downside of shRNA silencing constructs is the time and effort in producing individual shRNA plasmids and producing lentivirally integrated cell lines, although these processes can be automated to some extent.

Both siRNA and shRNA silences genes through RISC-complex-mediated degradation of targeted mRNA, a process which is dynamic and reversible. This requires a continuous presence of siRNA or shRNA in order to keep mRNA levels low, and a possible minor leakage from an inducible system will have negligible effects. CRIPSR/Cas9-mediated silencing on the other hand is an irreversible process. In our CRISPR/Cas9-compatible plasmid system we use a one-site target, often targeted to the first exon of a mRNA, that when damaged and repaired via NHEJ leads to the loss of a small number of basepairs and a frameshift of the mRNA sequence. Because it is irreversible, a small leakage of an inducible system would inevitably lead to a buildup of edited cells, provided the mutation is not lethal or affecting cellular growth. Therefore, our system combines an inducible Cas9 protein with inducible gRNA expression, to minimize this risk logarithmically.

5.2.4 Why Flexibility is an Important Feature of a Plasmid System

In a modern biology or biomedical laboratory, genetic manipulation of biological models is performed routinely, either via small fragments of DNA such as siRNA using chemical vectors, or in more stable systems such as viral delivery of new genetic material for either transient expression or stable integration in the cell's DNA. There are hundreds of available commercial

and non-commercial genetic vectors for these purposes, and many laboratories use many different systems.

In a biology or life science research setting, experimental designs are often prone to change due to unforeseen challenges and development of novel techniques. Throughout the life of a project, new experiments may be needed and results may necessitate the inclusion of experiments and tools not originally thought needed. In this dynamic setting, a flexible genetic engineering system can prove useful. Genetic tools developed for an experiment can be repurposed for another if the genetic vector system allows. This is the design principle of the vector system described in **Study IV**. Aside from a quick, modular and efficient way to move genetic codes for expression or downregulation to new vectors, the system also allows for rapid integration of new reporters and selection markers.

Utilizing a suite of fluorescent proteins over the entire range of visible light not only allows for a range of choices and combinations with reporter systems, but also multiplexing. Different genetic sequences can be experimentally combined after being subcloned into vectors with different fluorescent reporters, which allows for a high throughput and a higher degree of internal control in multi-fluorescent experiments

The plasmid system and laboratory predecessors described in **Study IV** has already been utilized by various laboratories, and while many collaborations are still ongoing, some have already resulted in published scientific articles (153–156) .

5.2.5 Designing a High Content Imaging Experiment – A Matter of Scale

Although the initial figures of **Paper III** describe results from a high-content imaging experiment comprised of ~16 000 multiplate wells, it is important to note that the experiment was not performed without prior optimization, both of individual reporter systems but also in the process of scaling up from smaller to higher throughput. Designing an experiment immediately on the highest throughput will undoubtedly lead to unforeseen problems and it is important to be aware of the necessity of pilot studies and optimization procedures. Thus, I wanted to briefly describe the process of the main experiment in **Study III** from the point of view from the design of the experiment, focusing on the path from initial experimental design to concluded experiment (**Figure 7**).

The initial experimental design was to combine four reporter systems in one experiment; Fucci, DAPI, EdU integration, and antibodies against 53bp1. Additionally, the experiment would be performed with two separate time points, and one set of the experiment would be irradiated prior to the first time point. Practically, each reporter system first had to be tested and optimized one by one in a plate format. Since the red and green fluorescent channel was already occupied by the integrated Fucci system, infrared and blue channels had to be utilized for EdU and 53bp1 reporters. As DAPI staining would be more than a hundred-fold higher intensity than an alexa-405-conjugated secondary antibody, it was decided to utilize the blue channel twice, first for the antibody and then after imaging restained with DAPI and imaged again. An alignment algorithm was used to merge the resulting images.

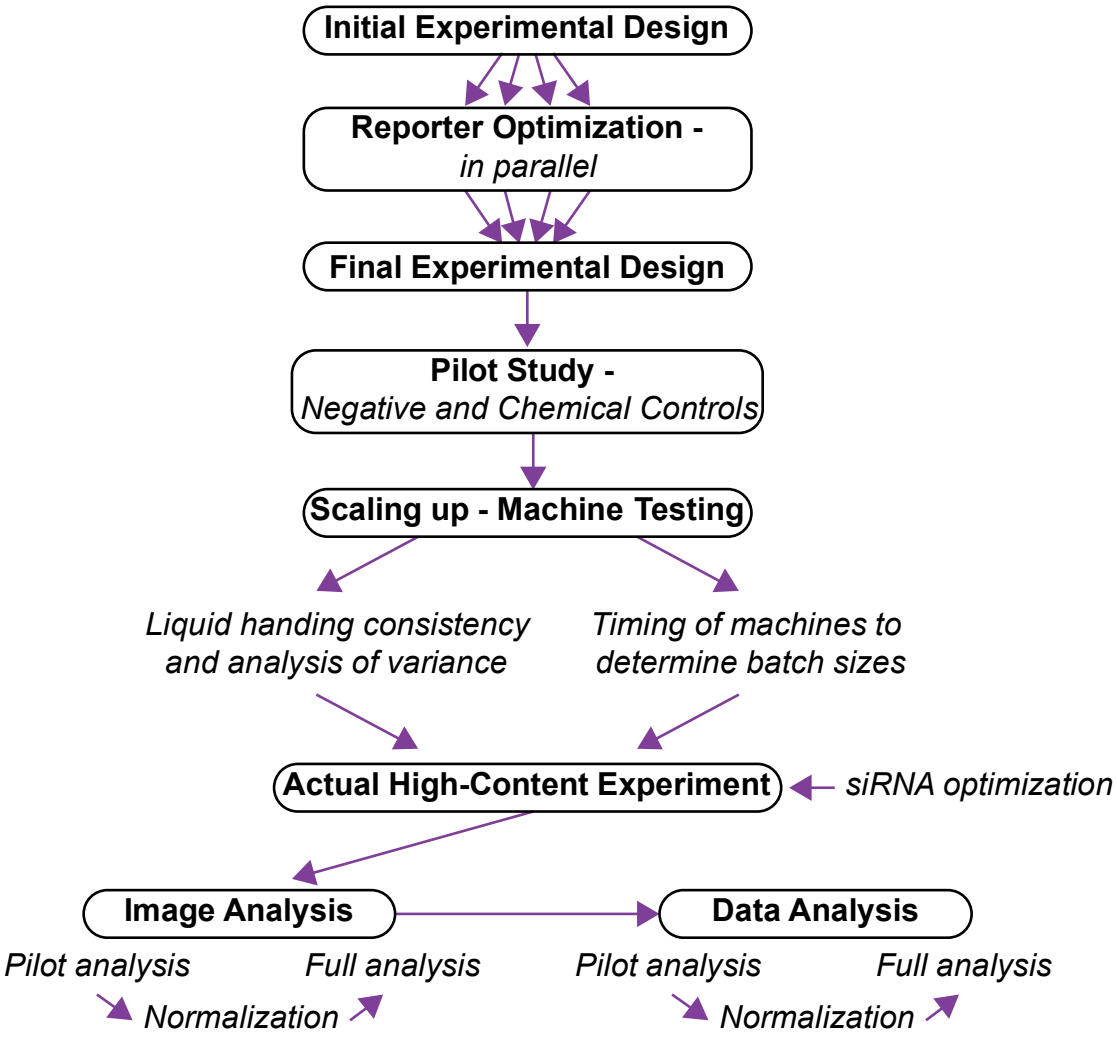

Figure 7. Design and optimization of a high-content experiment.

After all reporter systems were optimized and a protocol determined, we performed a pilot study on a series of chemicals with known cell cycle effects. During these trials it was discovered that the click-chemistry reaction (specifically, the cupric 2+ ions) had a quenching effect on the fluorescent proteins of the Fucci system. Fortunately, this quenching is reversible (157) and as the protocol continued with antibody incubations and a series of washes, the cupric ions were removed and fluorescence restored before final imaging.

After multiple test trials in 96-well format, including the addition of multiple time-points and external irradiation, the last optimization required scaling up to high-throughput 384-well format. This process focused on pre-emptively solving problems of scale. Scaling problems can be divided in timing and practical matters. Each handling of a plate takes a certain amount of time and it is easy to underestimate the combined handing times even for a simple washing step, possibly leading to differences in plate handling and unwanted addition of variance. Batch

sizes for each experimental procedure needed to be determined according to determined bottlenecks, and staggering steps introduced to the protocol to avoid downstream timing issues.

Practical problems of scale can be related to the small volumes and well sizes of the experiment. With smaller volumes, any variance in addition or removal of liquid can be a problem, and the consistency of plate aspirators needed to be measured, especially for small-volume removals such as antibody incubations. Foam is also a much larger problem in 384-well formats, and steps with detergent present might necessitate inclusion of a foam-removal step, such as gently dabbing the plate on a fiber-free tissue.

Even after an experiment is finished and all images are acquired, batch- or plate-specific effects can still appear and it is thus important to add enough negative controls for normalization after the experiment is finished. The addition of negative controls spread out over the plate can help identify sources of variance and aids in determining normalization procedures. When deciding on how controls can be included, it can be advantageous to wait until test trials are at the final scale, as scaling could reveal new methodological issues to control for. Different studies might necessitate different control distributions, be they a single column per plate, a semi-random placement over the plate, or a plate-per-plate randomization.

Image and data analysis after images are acquired normally follow the same iteration of developing analysis method for a specific feature, normalizing for any batch-wise or plate-wise effects, then designing a final analysis pipeline for use on the entire set. In order to keep this process objective and structured, it is advantageous to include chemical controls with known phenotypes in an image analysis study, as these control images can be utilized for development of the image and data analysis pipelines.

5.2.6 Data Analysis of High Content Imaging Data – a Structured Approach Compared to Computer-Assisted Methods

In modern medium and big data analysis pipelines, machine-assisted classifications are often utilized, either to assist or to replace defined categorical rules based on biology. While a human scientist can rely on their expertise in the field to draw up classification criteria, a machine can access a massive amount of information to guide classifications. Computer-assisted classifications has many advantages but lacks the intuition and experience of human scientists. It is a part of human psychology to insert new knowledge into older contexts, and while this can be perilous by risking confirmation bias or through overconfidence, it can be valuable to analyze biological and life science data in the context of prior biological knowledge. The utilization of intuition is dependent on objective expertise (158), and while it requires rigorous statistical tools to avoid confirmation biases, can be valuable in directing research.

During image analysis, especially in a high-content study, it is easy to generate an overwhelming amount of information, with many pieces of information being redundant to each other. When extracting feature descriptors in an image analysis pipeline, hundreds of features can easily be acquired from a simple experiment. Depending on the choice of analysis

method, be it based on strict biological criteria or through machine-learning-assisted classifiers, feature extraction from image data will inevitably vary.

For example, the question: “How bright is this cell”, can be answered either with the average pixel intensity, the integrated pixel intensity, the maximum pixel intensity, the average of the brightest 25% of pixels, or with even more advanced feature descriptors. Although many of these answers are highly correlated to each other, the choice of descriptor depends on the behavior of the reporter, background noise, prior segmentation problems, and the biological question. In the main siRNA experiment in **Study III** this is highlighted where the chemical EdU is measured with average intensity, the chromatin quantification measured with integrated intensity, and the two Fucci colors measured with upper quartile intensities. The choice was guided based on pilot analyses and biological interpretation, in order to choose the feature descriptor most likely to correspond to the biological question without being affected by methodological problems.

An alternative approach is to extract all possible information and use a machine learning tool to analyze and categorize cells. For human-assisted analysis, the inclusion of hundreds of variables can be confusing, but a computer has a larger accessible memory. If many of the features are highly correlated, a machine learning approach will have no problem taking this into account, and the combined information can sometimes be more valuable than a strategically chosen individual variable. In the validation of MAT2A’s cell-cycle dependent localization changes in **Study II**, we utilized the Classifier tool from CellProfiler Analyst (159) using all fluorescent feature descriptors from the segmentation of cytosol and nuclei to identify a cell-cycle dependent shift in localization.

The coming decades will be challenging in combining information from human and machine classifiers, and if done correctly, could combine the advantages of both while avoiding the pitfalls of each.

5.2.7 A Novel Use of Dimensionality Reduction for Visualization of Three-Dimensional Data.

In **Study I**, **II**, and **III**, the aims and biological questions necessitated the development of novel methodologies for visualization and comparison of high-dimensional data. In **Study I** and **II**, a polar coordinate system was utilized to visualize the cell cycle expression of transcripts and proteins respectively, and in **Study III** the distribution between three cell cycle phases is reduced to a two-dimensional representation of the data without loss of information.

For expression data, the variable “logarithmic expression” is a quantitative variable with no bounds or limits. The fold change difference between two groups belong to the same variable type, there are no maximum fold change differences. On the other hand, when comparing three distribution variables, both the prevalence of a specific type of cell and the difference between two prevalences are also ratio quantitative values, but these are finite in both directions, and cannot be smaller than 0% or higher than 100%. Aside from the difference in statistical approaches when analyzing these two variables, visualization also requires different strategies.

Reducing three dimensions to two without any loss of information is axiomatically impossible. A preferable characteristic of an algorithm producing a visual graph is that the coordinates should be two-dimensional to print on paper, or shown on a computer screen. Both methods had to approach this challenge, albeit in different ways.

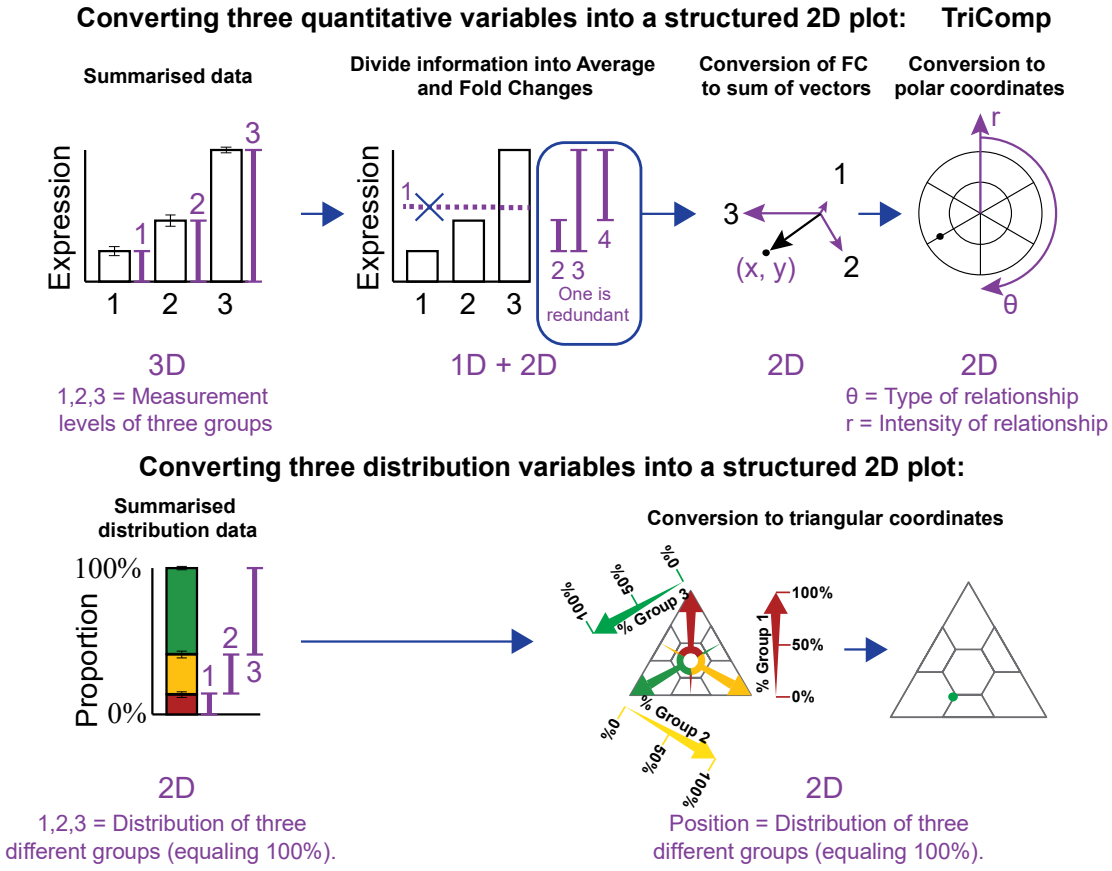

Figure 8. Reducing dimensions for visualization of the relationship between three groups.

For the polar coordinate system, we start with three variables; three measures of expression. If we do not mind losing the information regarding the average expression of the transcript or protein, if it is high- or low-expressed, we can express the relationship between the groups in two dimensions. The expression of the three groups can either be seen as three individual values, or it can be seen as an average value plus a relationship between three values (**Figure 8**). Such a relationship is two-dimensional, as the information about two of the differences is enough to provide the third, they thus carry redundancy. When treating it as three individual values we don't know which one to remove, as we deem all equally important for understanding the biology. But converting to polar coordinates allows us to plot the relationship between the three groups in two-dimensional space. An additional feature of the TriComp algorithm is that the two polar coordinates carry different information. The angle θ describes

the type of relationship and the radius r describes how intense the relationship is. This can be useful for comparing different experiments, by plotting θ-values against each other and filter on r and an accompanying p-value. The average expression was abandoned before polar conversion but we can still utilize the average expression later in downstream analysis for filtering or plotting using opacity, shape, or size, but it is superfluous in describing the relationship between the three groups.

The use of the TriComp algorithm also enables a high resolution of describing a relationship between the three groups. The θ variable is quantitative and each value corresponds to an exact relationship. Two different mRNAs that are highly expressed in G2-phase might be different during the rest of the cell cycle, and including information about S-phase expression can elucidate a more informed knowledge about cell cycle oscillation patterns. While the high resolution of time points after synchronizing cells (10) achieves a higher temporal resolution, a high number of time-points makes it harder to summarize patterns of cell cycle fluctuations, and necessitates the use of pattern-finding algorithms.

The triangularization algorithm summarizes three-group distribution patterns in a similar manner to TriComp, but an important distinction is that as the distribution of three groups is functionally two-dimensional if they together equal 100%, the data can directly be positioned in two-dimensional space (**Figure 8**). A triangle was chosen for symmetry, with each corner representing a full arrest towards that cell cycle phase. In distribution analysis, changes in distributions does not need logarithmic representation, and the triangular coordinate system is thus linear.

When comparing two different data sets to each other, a structured approach can be advantageous. While modern clustering methods, such as t-sne or Principal Component Analysis (PCA), are accessible and easy to use in many softwares, clustering coordinates from t-sne or PCA is created anew for each data analysis. In order to compare results from two different sets of data, they would need to be employed to categorize phenotypes independently and then the resulting phenotypes can be compared as two categorical variables. The TriComp method allows for a structured approach, and the resulting quantitative variable θ was used throughout **Study I** and **II** when compared to external data and between TriComp datasets. Throughout the studies TriComp data was compared to or stratified over categorical descriptors such as membership of protein families of transcription factors (160), kinases (161), phosphatases (162), E2F targets (163) or GO Ontology terms, ordinal variables such as an index of essentiality (113), and quantitative data such as previously characterized cell-cycle fluctuations (10,11).

The primary limitation of TriComp and the triangular visualization of distributions is that they are limited to three measured groups. Three expression levels can be reduced to two dimensions by ignoring the average value, but four or a higher amount of groups cannot. However, for biological questions with three natural groups, the algorithms have proved their usefulness. Both methods developed made it possible to summarize results at high information density,

categorize phenotypes, and compare results between conditions and to external data (**Figure 9**).

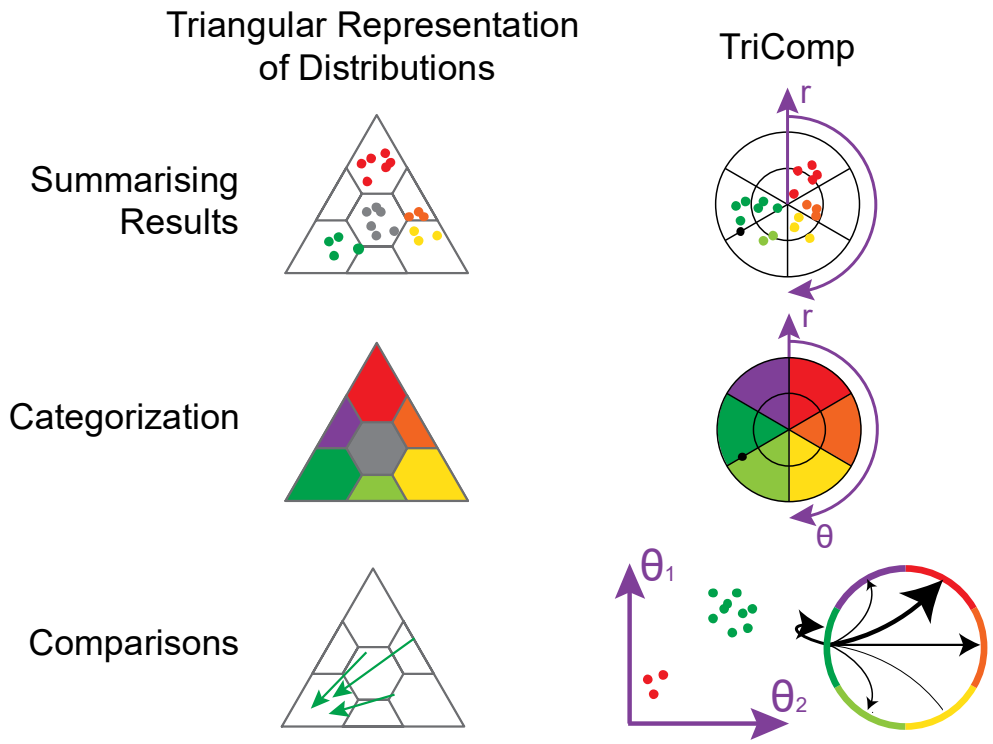

Figure 9. Utilizations of the TriComp and the triangularization algorithm.

5.2.8 Visualization of Multi-Parametric Data Using Geometry and Color

While paper or computer screen still images are limited to showing two dimensions, there are many additional ways to incorporate more dimensions in an image, such as color, opacity, shape, size, arrows, or integrated pie-graphs. See **Figure 10** for an early choice during Study III in summarizing all data on siRNA-silenced UPS genes. **Figure 10A** combines eight different phenotypical variables into a single donut-donut-circle-graphic, repeated for every UPS phenotype and clustered with a t-sne method. Fucci cell cycle distribution is represented as the outer ring in red/yellow/green, chromatin content distribution in the inner in blue-scales, the average EdU integration as the intensity of the inner circle, and viability as the size of the entire graphic. T-sne was used to cluster the data, which utilized the two coordinate dimensions. Note that while the image contains plenty of information, its usefulness to describe the results can be questioned. Too high density of information can quickly risk becoming cluttered and unintelligible. Reducing the number of included phenotypes results in a simpler more informative figure (**Figure 10B**), which only includes Fucci cell cycle distributions displayed as a pie graph, with sizes corresponding to viability. This results in a graph easier to interpret, confirming the proverb: “Less is more”. It is possible to see that most cell cycle arrests are

accompanied by a lower viability, and all siRNA pools not resulting in a phenotype are clustered some distance away.

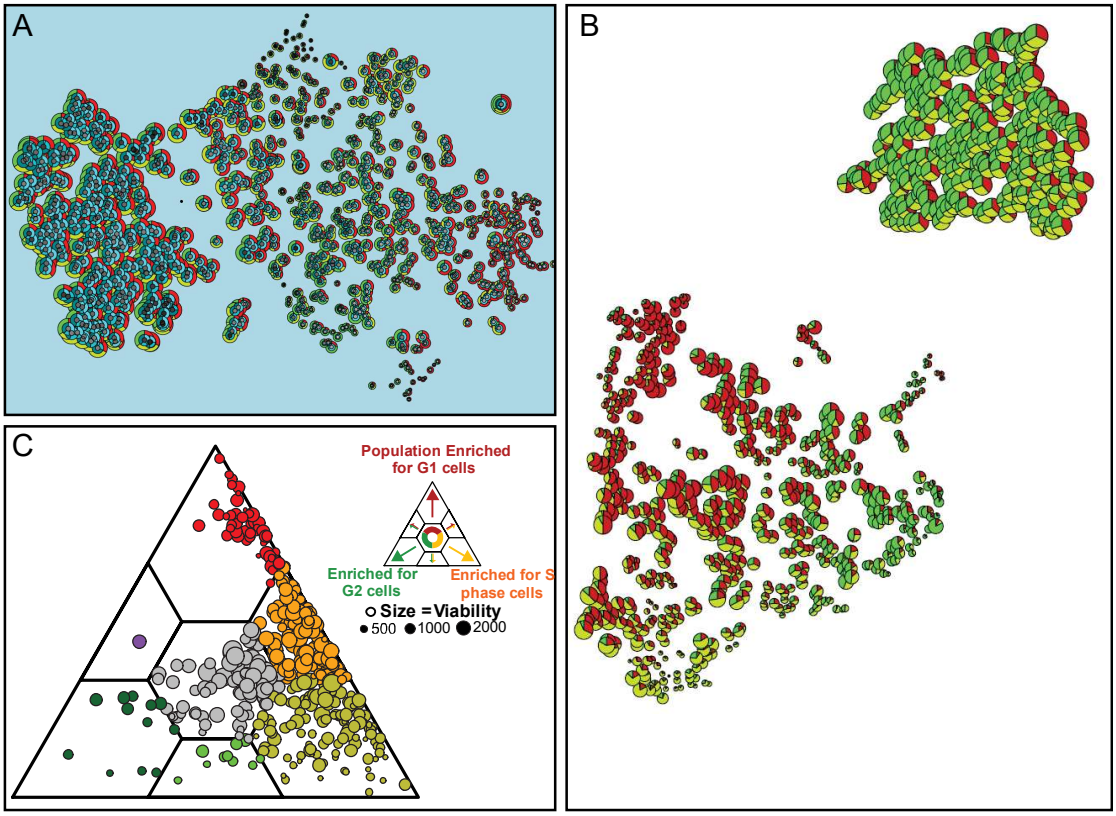

Fig 10. Three methods of clustering and visualizing cell cycle distribution phenotypes

The reason **Figure 10B** was never included in the manuscript is because it was unable to answer a central scientific question of the experiment; how cell cycle phenotypes are affected by time and irradiation. The resulting graphic needed to be able to answer whether the induced phenotype when silencing a gene would change depending on time and irradiation. In order to actually compare the changes visually, there arose a need for a structured coordinate system, defined as the same x-y-position carrying the same meaning between experiments. A structured triangular method of visualization was thus chosen (**Figure 10C**), containing the same information as the previous t-sne graphic. Although inspired from t-sne generated clustering, the major advantage with the structured approach is that the same point in space corresponds to an exact phenotypical distribution, and a specific distance in the graph corresponding to a specific change in cell cycle distribution patterns. The x-y-coordinates are used both to cluster and to describe the data, compared to t-sne methods which needs overlaid pie graphs to visualize the data. The structure also plotting allows for comparison of phenotypes over time and irradiation, by plotting arrows onto the triangular space (**Figure 11**).

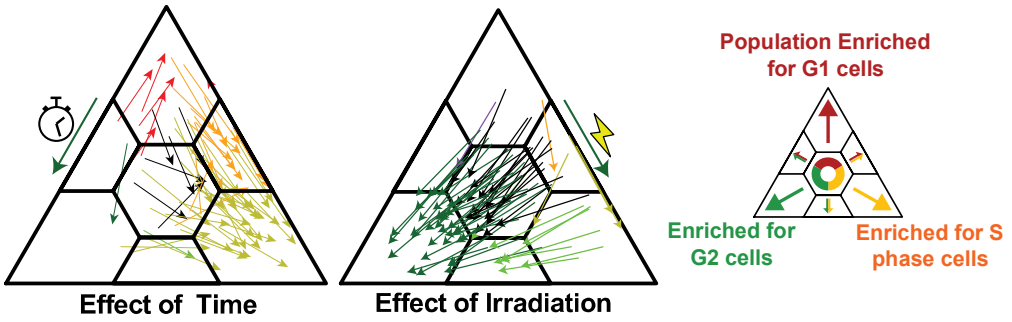

Figure 11. Comparison of cell cycle phenotypes over time and irradiation in a structured space

When introducing more dimensions through the use of colors, opacities, sizes, or shapes, it is important to avoid information overload, and presentation of multi-dimensional data is pedagogically improved if the biological information is inherently reflected in the graphic utilized. For example, size can be an useful indicator of cellular viability. Scaling the size of a graphic describing a cell cycle phenotype according to the viability serves multiple functions. Any viability phenotypes that are essentially lethal, i.e. no surviving cells, will minimize the graphic, the information about cell cycle distribution is less important if most of the cells are already dead. Additionally, the reader of the graph can easily internalize that size is an indicator of the amount of cells.

Opacity can be utilized without informational overload when filtering on statistical certainty, as its main effect on a graphic is hiding it, reducing its visibility. Similarly, shapes and colors are more appropriately used to differentiate between groups or subdivisions of data, perhaps to showcase differences or highlight a categorization based on a quantitative variable. Arrows are a useful graphical tool when visualizing paired data, although focus will inevitably be drawn to the longest arrows, making them useful when a change in coordinates correspond to shifts in searched for results.

A majority of the graphics produced throughout the studies comprising this thesis were produced with *ggplot2* (98), a package for the programming software R built under the design principle of enabling a "grammar of graphics", compared to most graph softwares' template-based designs. When designing custom visualization methods including the tools described above, *ggplot2* is heavily recommended by the author.

6 CONCLUDING REMARKS

Essential biological functions are often complex, and the cell cycle, DDR, and the UPS are no exception. Throughout this thesis, regulation and dysregulation of and by these processes were characterized and summarized. The cell cycle was found to affect a large amount of the cells' mRNA and protein, as well as functional processes such as phosphorylation and subcellular localizations. The UPS was functionally characterized regarding basic cellular processes, and a wide variety of phenotypes were established, mirroring the complexity of the family with over a thousand associated members.

Aside from providing results in resource formats, the studies also investigated novel roles of proteins regarding these processes. MAT2A was identified to change subcellular localization over the cell cycle, ARIH1 and ARIH2 were identified to be essential for GBM growth and proliferation, and the E3 adapter BTBD1 was found to be essential for viability, normal cell cycle progression, and a functional NHEJ-response to DNA double-strand breaks. Investigating the role of BTBD1 further, it was discovered to mirror certain aspects of the cytostatic drug Camptothecin and arrest TOP1 bound to DNA.

Both TriComp and the triangularization algorithm employed in these studies were developed in response to scientific questions which standard methodologies failed to answer. While there exists a multitude of methods of clustering and categorizing phenotypical data, a structured approach is advantageous to be able to compare results and link the presentation of data directly to biological information. By utilizing novel approaches in dimension transformation and selection of dimensions, a structured approach was enabled throughout the studies. Two distinct but similar algorithms were utilized to visualize, categorize, and compare mid-dimensional data in an informative manner. Although modern computer-assisted classifiers have no need for visualization, and can effortlessly juggle thousands of dimensions, the utilization of the expertise and intuition of human scientists requires presentation and visual inspection when presenting and analyzing biological and life science data. Humans are experts at finding patterns and correlating disparate pieces of information, but at higher complexities of information we struggle. I believe an efficient approach in high-throughput and high-content analysis is to reduce information overload through the use of structured approaches of dimensionality reduction where the biological information is visible in the final variables, combined with sound statistics and scientific methodology.

In this thesis, observational studies of the cycle transcriptomics and proteomics are combined with a systematic characterization of phenotypes of cellular processes upon removal of UPS proteins through genetic interference. By studying both the intrinsic regulation of and by the cell cycle through observational studies, and the dysregulation through the functional removal of essential proteins, an enhanced understanding of basic cellular processes can be acquired, as well as identifying novel protein candidates involved in critical intra-cellular processes.

7 POPULAR SCIENCE SUMMARY

Our bodies are built out of tiny compartments called cells, each with a different role to play to support its larger body, the organism. Every function in the organism is on a microscopic level carried out by these cells, and the process through which the cells multiply and grow is called The Cell Cycle, which is divided into distinct and carefully orchestrated phases:

A cell is born.	*Start*	
A cell waits for a signal	*G1 Phase*	
A cell copies its genetic material.	*S Phase*	
A cell grows and prepares to split into two.	*G2 Phase*	
A cell divides and becomes two daughter cells.	*Mitosis*	
A cell is born A cell is born.	*Restart*	

Inside the cell, three sets of molecules act together to control its functions. They are called DNA, mRNA and proteins. DNA is in charge of planning, and all its schematics have been inherited from previous generations. DNA is also a large molecule in need of protection from outside damage, and does not want to move far from its compartment in the middle of the cell. When the cell requires a specific schematic for execution, it copies its plan to mRNA molecules for transport to the building sites. At the building sites the proteins are being built according to the plans. Proteins are the molecules with "the real jobs" in the cell and each protein is built for a specific function; cutting things apart, gluing things back together, moving things back and forth, operating signals to other proteins, and some even take out the trash and clean up the cell.

This last group of proteins is quite interesting, as they are often thought to only discard unwanted proteins, or proteins that are old and degraded. In reality, there are over a thousand different versions of these proteins, and they use a quite complicated labeling system, where much of the labeling still has unknown purposes. The labeling uses a simple tag, called ubiquitin, which can be applied in many different ways with different meanings. The entire group of proteins are named based on their labeling system, and they are called "The Ubiquitin Proteasome System". It is now known that this labeling can be used for much more than just taking out the trash, and these proteins have been shown important for many processes in the cell, including control of the cell cycle and repair of DNA.

This thesis investigates the interplay between the RNA runners and the Protein workers (especially the ones taking out the trash), all in the context of the cell cycle and DNA repair. It further investigates how the cells' requirements change throughout the cell cycle and how the different systems rely on and affect each other.

In **Study I**, the RNA is investigated throughout the cell cycle, and a specific group of proteins in charge of deciding when the RNAs are created are focused upon.

In **Study II**, the focus shifts to the proteins, and investigates how the changes on RNA investigated in **Study I** translates to changes in proteins during the cell cycle. Protein functions are also investigated through the study of their modifications and where in the cell they are located.

In **Study III** the different kinds of ubiquitin-labeling proteins are studied by removing them one by one and investigating effects on the cell cycle, as well as on other basic processes in the cell. This is done by taking millions of photographs of cells, and letting a computer analyze each cell. With enough cells photographed, we can describe exactly where the cells are in the cell cycle and see which proteins were necessary for normal progression.

Study IV describes a laboratory system of manipulating the DNA in order to affect the amounts of RNA and proteins in the cell. This system can be used to study many different questions in biomedicine and biology, including evaluating two ubiquitin-labeling proteins on their effect on growth and cell cycle progression in multiple cell lines of glioblastoma multiforme, an aggressive form of brain cancer.

Throughout the studies, large and complex amounts of data were generated. Novel methods in visualization were developed and used to describe and analyze the resulting data. A common design principle of these methods aimed to pedagogically describe complex data through the use of geometry and colors, with the final visualized data reflecting the biology studied in the projects. This was used throughout the studies in order to distill meaningful information from complex data, which could be used to identify novel genes and proteins involved in basic cellular processes.

8 ACKNOWLEDGEMENTS

Very few things can be performed in solitude, and I want to extend my sincerest gratitude to all the people in my life whom without this thesis would have been impossible.

I want to first thank my **supervisors**, for guiding me through the PhD process from day one, answering all of my questions, and always being there for me.

Mikael Altun, my primary supervisor. Your methodological expertise, scientific rigor and ingenuity has been an inspiration through this process. You have sat with me on the floor pipetting bacteria, taught me how to haggle with vendors, and discussed countless scientific projects with me. You have shown me how open collaborative structures can lead to fruitful research, and you have often challenged me outside my comfort zone. I have grown in my scientific role through your supervision and mentorship.

Michael Andäng, a master in conceptualizing research and connecting dots between biological processes to create intricate thought-provoking patterns. Discussing research with you has never been boring.

Thomas Helleday, thank you for introducing me to the field of translational medicine and DNA repair, and the unique research lab environment you set up and let me work in.

Eric Rullman, thank you for all the discussions about algorithms, data analysis and statistics, for helping me formalize my methods, and for introducing me to machine learning concepts.

All my **colleagues** at my two primary places of work, both at the **Science for Life Laboratory** in Solna, and at **ANA Futura** in Huddinge. Thank you for all the support, fruitful discussions, and joyful distractions. Especially my closest colleagues, **Carolyn**, **Magdalena**, **Patrick**, **Saeed,** and **Andreas** at SciLifeLab, and **Nicholas**, **Seher** and **Amarjit** at ANA Futura. I also want to thank everyone else at the **Thomas Helleday Lab** and in the **Thomas Gustafsson Group** at ANA Futura. I've felt at home in both since the first day and I have been blessed to have such good people around me. I am sorry about all the pens I've stolen from you throughout the years though!

All my **colleagues** at collaborating laboratories around Karolinska Institutet and abroad; **Yaser**, **Gözde** and **Aditya** at the **Julian Walfridsson** Group, **Helena**, **Petra**, and **Zuzana** at the **Michael Andäng** Group, **Dimitris** at the **Jiri Bartek** group, **Gianluca** at the **Mattias Uhlen** Group, **Jing** at the **Simon Elsässer** Lab, the **Vítězslav Bryja** Group and the **Sten Linnarsson** Group. Without such an extensive collaborative support network, none of this would have been possible.

Research is neither possible without funding and I especially want to thank the **Hållsten foundation** for their generous support throughout the years.

An accessible national scientific infrastructure has also been critical for these studies, and has taught me the usefulness of centralizing expertise in complex methodologies. The following

laboratories and facilities have been critical in overcoming challenges throughout these projects and the members of each has been extremely helpful; the **Laboratories for Chemical Biology at Karolinska Institutet** (LCBKI) for advising on assay development and access to high-content imaging instruments, **Johan Viklund** at the **Bioinformatics Infrastructure for Life Sciences** (BILS) for support related to SQL programming in R, the **National Genomics Infrastructure** (NGI) for access to RNA Sequencing, the **CLICK Imaging Facility** for access to a high-content imaging microscope, and the **Uppsala Multidisciplinary Center for Advanced Computational Science** (UPPMAX) for providing computational resources for image analysis.

I want to give an extra thanks to **Bo**, **Ulf**, and **Maria** at the **Biochemical and Cellular Assay Facility** at SciLifeLab, for the close collaboration while scaling up image analysis experiments, and to **Carolina Wählby** at the **SciLifeLab BioImage Informatics Facility**, for patiently teaching me image analysis methods, taking the time to discuss my projects, and for the incredibly useful advice to learn the R programming software.

I also want to thank everyone else at Karolinska who has been integral for this journey, **Alessandra**, **Arja** and **Ann** for always reminding me about my yearly reports and answering all my questions about the PhD process. **Kristina**, **Sabina**, and **Moa** for helping me with just about everything!

I also want to thank my **family**, a source of foundation, support and replenishment of mental batteries. My **parents**, who have always been an academic inspiration to me. **Ola,** throughout my life you have always inspired and challenged my curiosity through problem solving and scientific discussions. **Maria**, your academic skill, rigor and integrity is exemplary and something I will always strive towards in my scientific career. My siblings, **Arvid**, **Anna** and **Maja**. With you, growing up was never lonely, and growing older will not be either.

All my fantastic **friends** in Stockholm and elsewhere I want to thank for all the adventures, events, discussions and support, I am truly honored to have all of you as friends. **Miloš**, you left Lund a year before me and I've followed in your footsteps since, thank you for paving the way! **Karen** and **Johan**, our almost-annual ten days of isolation in the wilderness has been critical for recharging batteries and enjoying nature, I hope we will go on many more trips. **Sanja**, my dear friend, thank you for always being there to support me, and for all the joyful times we've had. **Stelia**, **Eric**, **Žiga**, **Laura**, **Hannes**, **Martina**, **Ludde**, **Sofia** and everyone else around me, I can't wait for the current events to be over so we can hang out properly again, and party like before. My old friends from **Unga Forskare**, especially **Beatrice**, **Gustav**, and **Harald**, for countless scientific discussions and plenty of fun too.

Finally, I want to thank all my **teachers**, from childhood to adulthood. For enabling my curiosity, challenging me, and always pushing me forward.

9 REFERENCES

1. Bertoli C, Skotheim JM, de Bruin RAM. Control of cell cycle transcription during G1 and S phases. Nat Rev Mol Cell Biol. 2013 Jul 23;14(8):518–28.

2. Malumbres M, Barbacid M. Cell cycle, CDKs and cancer: a changing paradigm. Nat Rev Cancer. 2009 Mar;9(3):153–66.

3. Davis PK, Ho A, Dowdy SF. Biological methods for cell-cycle synchronization of mammalian cells. Vol. 30, BioTechniques. Eaton Publishing Company; 2001. p. 1322–31.

4. Langan TJ, Rodgers KR, Chou RC. Synchronization of mammalian cell cultures by serum deprivation. In: Methods in Molecular Biology. Humana Press Inc.; 2016. p. 97–105.

5. Eliassen KA, Baldwin A, Sikorski EM, Hurt MM. Role for a YY1-Binding Element in Replication-Dependent Mouse Histone Gene Expression. Mol Cell Biol. 1998 Dec 1;18(12):7106–18.

6. Sakaue-Sawano A, Kurokawa H, Morimura T, Hanyu A, Hama H, Osawa H, et al. Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. Cell. 2008 Feb 8;132(3):487–98.

7. Daynac M, Morizur L, Kortulewski T, Gauthier LR, Ruat M, Mouthon MA, et al. Cell sorting of neural stem and progenitor cells from the adult mouse subventricular zone and live-imaging of their cell cycle dynamics. J Vis Exp. 2015 Sep 14;2015(103).

8. Mort RL, Ford MJ, Sakaue-Sawano A, Lindstrom NO, Casadio A, Douglas AT, et al. Fucci2a: A bicistronic cell cycle reporter that allows Cre mediated tissue specific expression in mice. Cell Cycle. 2014 Sep 1;13(17):2681–96.

9. Pauklin S, Vallier L. The cell-cycle state of stem cells determines cell fate propensity. Cell. 2013 Sep 26;155(1):135.

10. Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, et al. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol Biol Cell. 2002 Jun;13(6):1977–2000.

11. Dominguez D, Tsai Y-H, Gomez N, Jha DK, Davis I, Wang Z. A high-resolution transcriptome map of cell cycle reveals novel connections between periodic genes and cancer. Cell Res. 2016 Aug 1;26(8):946–62.

12. Edfors F, Danielsson F, Hallström BM, Käll L, Lundberg E, Pontén F, et al. Gene-specific correlation of RNA and protein levels in human cells and tissues. Mol Syst Biol. 2016 Oct;12(10):883.

13. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet. 2012 Apr 13;13(4):227–32.

14. Liu Y, Beyer A, Aebersold R. On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell. 2016 Apr 21;165(3):535–50.

15. Orre LM, Vesterlund M, Pan Y, Arslan T, Zhu Y, Fernandes Woodbridge A, et al. SubCellBarCode: Proteome-wide mapping of protein localization and relocalization. Mol Cell.

16. Ly T, Ahmad Y, Shlien A, Soroka D, Mills A, Emanuele MJ, et al. A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells. Elife. 2014 Mar 4;3.

17. Ly T, Whigham A, Clarke R, Brenes-Murillo AJ, Estes B, Madhessian D, et al. Proteomic analysis of cell cycle progression in asynchronous cultures, including mitotic subphases, using PRIMMUS. Elife. 2017 Oct 20;6.

18. Dephoure N, Zhou C, Villén J, Beausoleil SA, Bakalarski CE, Elledge SJ, et al. A quantitative atlas of mitotic phosphorylation. Proc Natl Acad Sci U S A. 2008 Aug;105(31):10762–7.

19. Olsen J V, Vermeulen M, Santamaria A, Kumar C, Miller ML, Jensen LJ, et al. Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci Signal. 2010 Jan;3(104):ra3--ra3.

20. Becher I, Andrés-Pons A, Romanov N, Stein F, Schramm M, Baudin F, et al. Pervasive Protein Thermal Stability Variation during the Cell Cycle. Cell. 2018 Apr;173(6):1–32.

21. Dai L, Zhao T, Bisteau X, Sun W, Prabhu N, Lim YT, et al. Modulation of Protein-Interaction States through the Cell Cycle. Cell. 2018 Apr;173(6):1–28.

22. Wilkinson KD. Ubiquitination and deubiquitination: Targeting of proteins for degradation by the proteasome. Semin Cell Dev Biol. 2000 Jun;11(3):141–8.

23. Cappadocia L, Lima CD. Ubiquitin-like Protein Conjugation: Structures, Chemistry, and Mechanism. Chem Rev. 2017 Feb 24;acs.chemrev.6b00737.

24. Nayak A, Müller S. SUMO-specific proteases/isopeptidases: SENPs and beyond.

25. Ardley HC, Robinson PA. E3 ubiquitin ligases.

26. Genschik P, Sumara I, Lechner E. The emerging family of CULLIN3-RING ubiquitin ligases (CRL3s): cellular functions and disease implications. EMBO J. 2013 Aug 28;32(17):2307–20.

27. Carim-Todd L, Sumoy L, Andreu N, Estivill X, Escarceller M. Identification and characterization of BTBD1, a novel BTB domain containing gene on human chromosome 15q24. Gene. 2001 Jan;262(1–2):275–81.

28. Dhanoa BS, Cogliati T, Satish AG, Bruford EA, Friedman JS. Update on the Kelch-like (KLHL) gene family. Hum Genomics. 2013 May 15;7(1):13.

29. BioGrid. Cul3-BTBD1 interaction Summary [Internet]. 2017 [cited 2017 Mar 22]. Available from: https://thebiogrid.org/interaction/1219119/btbd1-cul3.html

30. Xu L, Yang L, Hashimoto K, Anderson M, Kohlhagen G, Pommier Y, et al. Characterization of BTBD1 and BTBD2, two similar BTB-domain-containing Kelch-like proteins that interact with Topoisomerase I. BMC Genomics. 2002;3:1.

31. Wu W, Hodges E, Höög C. Thorough validation of siRNA-induced cell death phenotypes defines new anti-apoptotic protein. Nucleic Acids Res. 2006 Jan 30;34(2):e13–e13.

32. Koegl M, Hoppe T, Schlenker S, Ulrich HD, Mayer TU, Jentsch S. A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly. Cell. 1999 Mar 5;96(5):635–44.

33. Akutsu M, Dikic I, Bremm A. Ubiquitin chain diversity at a glance. J Cell Sci. 2016 Mar 1;129(5):875–80.

34. Matsumoto ML, Wickliffe KE, Dong KC, Yu C, Bosanac I, Bustos D, et al. K11-linked polyubiquitination in cell cycle control revealed by a K11 linkage-specific antibody. Mol Cell. 2010 Aug;39(3):477–84.

35. Min M, Mevissen TET, De Luca M, Komander D, Lindon C. Effcient APC/C substrate degradation in cells undergoing mitotic exit depends on K11 ubiquitin linkages. Mol Biol Cell. 2015 Dec 1;26(24):4325–32.

36. Gatti M, Pinato S, Maiolica A, Rocchio F, Prato MG, Aebersold R, et al. RNF168 promotes noncanonical K27ubiquitination to signal DNA damage. Cell Rep. 2015 Jan 13;10(2):226–38.

37. Nijman SMB, Luna-Vargas MP a, Velds A, Brummelkamp TR, Dirac AMG, Sixma TK, et al. A genomic and functional inventory of deubiquitinating enzymes. Cell. 2005 Dec 2;123(5):773–86.

38. Amerik AY, Hochstrasser M. Mechanism and function of deubiquitinating enzymes. Biochim Biophys Acta - Mol Cell Res. 2004 Nov;1695(1–3):189–207.

39. Upadhya SC, Hegde AN. Role of the ubiquitin proteasome system in Alzheimer's disease. BMC Biochem. 2007 Jan;8 Suppl 1(Suppl 1):S12.

40. Shangary S, Wang S. Targeting the MDM2-p53 interaction for cancer therapy. Clin Cancer Res. 2008 Sep 1;14(17):5318–24.

41. Crawford LJ, Walker B, Irvine AE. Proteasome inhibitors in cancer therapy. J Cell Commun Signal. 2011 Jun;5(2):101–10.

42. D'Arcy P, Brnjic S, Olofsson MH, Fryknäs M, Lindsten K, De Cesare M, et al. Inhibition of proteasome deubiquitinating activity as a new cancer therapy. Nat Med. 2011 Nov 6;17(12):1636–40.

43. Wang X, Mazurkiewicz M, Hillert E-K, Olofsson MH, Pierrou S, Hillertz P, et al. The proteasome deubiquitinase inhibitor VLX1570 shows selectivity for ubiquitin-specific protease-14 and induces apoptosis of multiple myeloma cells. Sci Rep. 2016 Jul 6;6(1):26979.

44. Sowa ME, Bennett EJ, Gygi SP, Harper JW. Defining the human deubiquitinating enzyme interaction landscape. Cell. 2009 Jul 23;138(2):389–403.

45. Kim W, Bennett EJ, Huttlin EL, Guo A, Li J, Possemato A, et al. Systematic and quantitative assessment of the ubiquitin-modified proteome. Mol Cell. 2011 Oct 21;44(2):325–40.

46. Zhang X, Smits AH, van Tilburg GBA, Jansen PWTC, Makowski MM, Ovaa H, et al. An Interaction Landscape of Ubiquitin Signaling. Mol Cell. 2017 Mar 2;65(5):941-955.e8.

47. Jackson SP, Bartek J. The DNA-damage response in human biology and disease. Vol. 461, Nature. Nature; 2009. p. 1071–8.

48. Helleday T, Petermann E, Lundin C, Hodgson B, Sharma RA. DNA repair pathways as targets for cancer therapy. Nat Rev Cancer. 2008 Mar;8(3):193–204.

49. Houtgraaf JH, Versmissen J, van der Giessen WJ. A concise review of DNA damage checkpoints and repair in mammalian cells. Cardiovasc Revascularization Med. 2006 Jul;7(3):165–72.

50. Brinkmann K, Schell M, Hoppe T, Kashkar H. Regulation of the DNA damage response by ubiquitin conjugation. Front Genet. 2015;6(MAR):1–15.

51. Jacq X, Kemp M, Martin NMB, Jackson SP. Deubiquitylating Enzymes and DNA Damage Response Pathways. Cell Biochem Biophys. 2013 May 28;

52. Falaschetti CA, Mirkin EC, Raha S, Paunesku T, Woloschak GE. The Ubiquitin-Proteasome System and DNA Repair. In: DNA Repair - On the Pathways to Fixing DNA Damage and Errors. InTech; 2011.

53. Li W, Nagaraja S, Delcuve GP, Hendzel MJ, Davie JR. Effects of histone acetylation, ubiquitination and variants on nucleosome stability. Biochem J. 1993;296(3):737–44.

54. Park JM, Yang SW, Yu KR, Ka SH, Lee SW, Seol JH, et al. Modification of PCNA by ISG15 Plays a Crucial Role in Termination of Error-Prone Translesion DNA Synthesis. Mol Cell. 2014 May;54(4):626–38.

55. Jeon YJ, Park JH, Chung CH. Interferon-Stimulated Gene 15 in the Control of Cellular Responses to Genotoxic Stress. Mol Cells. 2017 Feb 28;40(2):83–9.

56. San Filippo J, Sung P, Klein H. Mechanism of eukaryotic homologous recombination. Vol. 77, Annual Review of Biochemistry. Annu Rev Biochem; 2008. p. 229–57.

57. Lieber MR. The mechanism of human nonhomologous DNA End joining. Vol. 283, Journal of Biological Chemistry. J Biol Chem; 2008. p. 1–5.

58. Kakarougkas A, Jeggo PA. DNA DSB repair pathway choice: An orchestrated handover mechanism. Vol. 87, British Journal of Radiology. British Institute of Radiology; 2014.

59. Symington LS, Gautier J. Double-strand break end resection and repair pathway choice. Annu Rev Genet. 2011;45:247–71.

60. Ha K, Ma C, Lin H, Tang L, Lian Z, Zhao F, et al. The anaphase promoting complex impacts repair choice by protecting ubiquitin signalling at DNA damage sites. Nat Commun. 2017 Jun 12;8(1):1–17.

61. Doil C, Mailand N, Bekker-Jensen S, Menard P, Larsen DH, Pepperkok R, et al. RNF168 Binds and Amplifies Ubiquitin Conjugates on Damaged Chromosomes to Allow Accumulation of Repair Proteins. Cell. 2009 Feb 6;136(3):435–46.

62. Mailand N, Bekker-Jensen S, Faustrup H, Melander F, Bartek J, Lukas C, et al. RNF8 Ubiquitylates Histones at DNA Double-Strand Breaks and Promotes Assembly of Repair Proteins. Cell. 2007 Nov 30;131(5):887–900.

63. Bartocci C, Denchi EL. Put a RING on it: regulation and inhibition of RNF8 and RNF168 RING finger E3 ligases at DNA damage sites. Front Genet. 2013;4:128.

64. Oda H, Hübner MR, Beck DB, Vermeulen M, Hurwitz J, Spector DL, et al. Regulation of the Histone H4 Monomethylase PR-Set7 by CRL4Cdt2-Mediated PCNA-Dependent Degradation during DNA Damage. Mol Cell. 2010 Nov;40(3):364–76.

65. Zhang H, Liu H, Chen Y, Yang X, Wang P, Liu T, et al. A cell cycle-dependent BRCA1-UHRF1 cascade regulates DNA double-strand break repair pathway choice. Nat

Commun. 2016 Jan;7:10201.

66. Schwertman P, Bekker-Jensen S, Mailand N. Regulation of DNA double-strand break repair by ubiquitin and ubiquitin-like modifiers. Nat Rev Mol Cell Biol. 2016 May 23;17(6):379–94.

67. Nakada S, Tai I, Panier S, Al-Hakim A, Iemura SI, Juang YC, et al. Non-canonical inhibition of DNA damage-dependent ubiquitination by OTUB1. Nature. 2010 Aug 19;466(7309):941–6.

68. Murai J, Yang K, Dejsuphong D, Hirota K, Takeda S, D'andrea AD. The USP1/UAF1 Complex Promotes Double-Strand Break Repair through Homologous Recombination †. Mol Cell Biol. 2011;31(12):2462–9.

69. Bassermann F, Eichner R, Pagano M. The ubiquitin proteasome system - Implications for cell cycle control and the targeted treatment of cancer. Vol. 1843, Biochimica et Biophysica Acta - Molecular Cell Research. NIH Public Access; 2014. p. 150–62.

70. Vodermaier HC. APC/C and SCF: Controlling each other and the cell cycle. Vol. 14, Current Biology. Cell Press; 2004. p. R787–96.

71. Nakayama KI, Nakayama K. Ubiquitin ligases: cell-cycle control and cancer. Nat Rev Cancer. 2006 May;6(5):369–81.

72. Zheng N, Wang Z, Wei W. Ubiquitination-mediated degradation of cell cycle-related proteins by F-box proteins. Int J Biochem Cell Biol. 2016 Feb 6;

73. Kim Y, Kipreos ET. Cdt1 degradation to prevent DNA re-replication: conserved and non-conserved pathways. Cell Div. 2007 Jan;2(1):18.

74. Jones HW, McKusick VA, Harper PS, Wuu KD. George Otto Gey. (1899-1970). The HeLa cell and a reappraisal of its origin. Obstet Gynecol. 1971 Dec 1;38(6):945–9.

75. Ponten J, Saksela E. Two established in vitro cell lines from human mesenchymal tumours. Int J Cancer. 1967;2(5):434–47.

76. Pudelko L, Edwards S, Balan M, Nyqvist D, Al-Saadi J, Dittmer J, et al. An orthotopic glioblastoma animal model suitable for high-throughput screenings. Neuro Oncol. 2018;20(11):1475–84.

77. Bajar BT, Lam AJ, Badiee RK, Oh YH, Chu J, Zhou XX, et al. Fluorescent indicators for simultaneous reporting of all four cell cycle phases. Vol. 13, Nature Methods. Nature Publishing Group; 2016. p. 993–6.

78. Dull T, Zufferey R, Kelly M, Mandel RJ, Nguyen M, Trono D, et al. A Third-Generation Lentivirus Vector with a Conditional Packaging System. J Virol. 1998 Nov 1;72(11):8463–71.

79. Katzen F. Gateway(®) recombinational cloning: a biological operating system. Expert Opin Drug Discov. 2007 Apr 14;2(4):571–89.

80. Reece-Hoyes JS, Walhout AJM. Gateway recombinational cloning. Vol. 2018, Cold Spring Harbor Protocols. Cold Spring Harbor Laboratory Press; 2018. p. 1–6.

81. Shaner NC, Steinbach PA, Tsien RY. A guide to choosing fluorescent proteins. Nat Methods. 2005 Dec;2(12):905–9.

82. Prendergast FG, Mann KG. Chemical and Physical Properties of Aequorin and the Green Fluorescent Protein Isolated from Aequorea forskålea†. Biochemistry. 1978;17(17):3448–53.

83. Goedhart J, Von Stetten D, Noirclerc-Savoye M, Lelimousin M, Joosen L, Hink MA, et al. Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. Nat Commun. 2012;3.

84. Ormö M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington J. Crystal structure of the Aequorea victoria green fluorescent protein. Science (80-). 1996 Sep 6;273(5280):1392–5.

85. Shaner NC, Campbell RE, Steinbach PA, Giepmans BNG, Palmer AE, Tsien RY. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. Nat Biotechnol. 2004 Dec;22(12):1567–72.

86. Subach OM, Cranfill PJ, Davidson MW, Verkhusha V V. An enhanced monomeric blue fluorescent protein with the high chemical stability of the chromophore. PLoS One. 2011 Dec 8;6(12).

87. Shcherbakova DM, Verkhusha V V. Near-infrared fluorescent proteins for multicolor in vivo imaging. Nat Methods. 2013 Aug;10(8):751–4.

88. Nguyen HB, Hung LW, Yeates TO, Terwilliger TC, Waldo GS. Split green fluorescent protein as a modular binding partner for protein crystallization. Acta Crystallogr Sect D Biol Crystallogr. 2013 Dec 19;69(12):2513–23.

89. Iwano S, Sugiyama M, Hama H, Watakabe A, Hasegawa N, Kuchimaru T, et al. Single-cell bioluminescence imaging of deep tissue in freely moving animals. Science (80-). 2018 Feb 23;359(6378):935–9.

90. Hall MP, Unch J, Binkowski BF, Valley MP, Butler BL, Wood MG, et al. Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. ACS Chem Biol. 2012 Nov 16;7(11):1848–57.

91. Lambert TJ. FPbase: a community-editable fluorescent protein database. Vol. 16, Nature Methods. Nature Publishing Group; 2019. p. 277–8.

92. Anoopkumar-Dukie S, Carey JB, Conere T, O'Sullivan E, Van Pelt FN, Allshire A. Resazurin assay of radiation response in cultured cells. Br J Radiol. 2005 Oct;78(934):945–7.

93. Chehrehasa F, Meedeniya ACB, Dwyer P, Abrahamsen G, Mackay-Sim A. EdU, a new thymidine analogue for labelling proliferating cells in the nervous system. J Neurosci Methods. 2009 Feb 15;177(1):122–30.

94. Himo F, Lovell T, Hilgraf R, Rostovtsev V V., Noodleman L, Sharpless KB, et al. Copper(I)-catalyzed synthesis of azoles. DFT study predicts unprecedented reactivity and intermediates. J Am Chem Soc. 2005 Jan 12;127(1):210–6.

95. Hong V, Presolski SI, Ma C, Finn MG. Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. Angew Chemie - Int Ed. 2009 Dec 21;48(52):9879–83.

96. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2017.

97. Wickham H, Francois R. dplyr: A Grammar of Data Manipulation. 2016.

98. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer; 2009.

99. Altelaar AFM, Frese CK, Preisinger C, Hennrich ML, Schram AW, Timmers HTM, et al. Benchmarking stable isotope labeling based quantitative proteomics. J Proteomics. 2013 Nov 12;88:14–26.

100. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008 Dec 30;26(12):1367–72.

101. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen J V., Mann M. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. J Proteome Res. 2011 Apr 1;10(4):1794–805.

102. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. Vol. 13, Nature Methods. Nature Publishing Group; 2016. p. 731–40.

103. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: Archive for functional genomics data sets - Update. Nucleic Acids Res. 2013 Jan 1;41(D1):D991–5.

104. Vizcaíno JA, Côté RG, Csordas A, Dianes JA, Fabregat A, Foster JM, et al. The Proteomics Identifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res. 2012 Nov 29;41(D1):D1063–9.

105. Kamentsky L, Jones TR, Fraser A, Bray MA, Logan DJ, Madden KL, et al. Improved structure, function and compatibility for cellprofiler: Modular high-throughput image analysis software. Bioinformatics. 2011 Apr;27(8):1179–80.

106. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, et al. CellProfiler: Image analysis software for identifying and quantifying cell phenotypes. Genome Biol. 2006 Oct 31;7(10):R100.

107. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010 Jan 1;26(1):139–40.

108. Engels B. XNomial: Exact Goodness-of-fit Test for Multinomial Data with Fixed Probabilities. 2015.

109. Carbon S, Douglass E, Dunn N, Good B, Harris NL, Lewis SE, et al. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 2019 Jan 8;47(D1):D330–8.

110. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: Tool for the unification of biology. Vol. 25, Nature Genetics. NIH Public Access; 2000. p. 25–9.

111. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37(1):1–13.

112. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large

gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44–57.

113. Bertomeu T, Coulombe-Huntington J, Chatr-Aryamontri A, Bourdages KG, Coyaud E, Raught B, et al. A High-Resolution Genome-Wide CRISPR/Cas9 Viability Screen Reveals Structural Features and Contextual Diversity of the Human Cell-Essential Proteome. Mol Cell Biol. 2018 Jan 1;38(1).

114. Medvar B, Raghuram V, Pisitkun T, Sarkar A, Knepper MA. Comprehensive database of human E3 ubiquitin ligases: application to aquaporin-2 regulation. Physiol Genomics. 2016;48(7):502–12.

115. Zhou J, Xu Y, Lin S, Guo Y, Deng W, Zhang Y, et al. iUUCD 2.0: an update with rich annotations for ubiquitin and ubiquitin-like conjugations. Nucleic Acids Res. 2018 Jan 4;46(D1):D447–53.

116. Li W, Bengtson MH, Ulbrich A, Matsuda A, Reddy VA, Orth A, et al. Genome-Wide and Functional Annotation of Human E3 Ubiquitin Ligases Identifies MULAN, a Mitochondrial E3 that Regulates the Organelle's Dynamics and Signaling. Ploegh H, editor. PLoS One. 2008 Jan 23;3(1):e1487.

117. Khurana B, Zhuang L, Moitra PK, Stantchev TS, Broder CC, Cutler M, et al. Human TOP1 residues implicated in species specificity of HIV-1 infection are required for interaction with BTBD2, and RNAi of BTBD2 in old world monkey and human cells increases permissiveness to HIV-1 infection. Virol J. 2010 Nov 20;7(1):332.

118. Campeau E, Ruhl VE, Rodier F, Smith CL, Rahmberg BL, Fuss JO, et al. A versatile viral system for expression and depletion of proteins in mammalian cells. PLoS One. 2009 Aug 6;4(8):6529.

119. Boström J, Sramkova Z, Salašová A, Johard H, Mahdessian D, Fedr R, et al. Comparative cell cycle transcriptomics reveals synchronization of developmental transcription factor networks in cancer cells. PLoS One. 2017 Dec;12(12):e0188772.

120. Wu L, Timmers C, Malti B, Saavedra HI, Sang L, Chong GT, et al. The E2F1-3 transcription factors are essential for cellular proliferation. Nature. 2001 Nov 22;414(6862):457–62.

121. La Thangue NB. DP and E2F proteins: components of a heterodimeric transcription factor implicated in cell cycle control. Curr Opin Cell Biol. 1994 Jun 1;6(3):443–50.

122. Polager S, Kalma Y, Berkovich E, Ginsberg D. E2fs up-regulate expression of genes involved in DNA replication, DNA repair and mitosis. Oncogene. 2002 Jan 17;21(3):437–46.

123. Muratovska A, Zhou C, He S, Goodyer P, Eccles MR. Paired-Box genes are frequently expressed in cancer and often required for cancer cell survival. Oncogene. 2003 Sep 11;22(39):7989–97.

124. Morgan R. Pax6 is a direct, positively regulated target of the circadian geneClock. Dev Dyn. 2004 Aug 1;230(4):643–50.

125. Hsieh YW, Yang XJ. Dynamic Pax6 expression during the neurogenic cell cycle influences proliferation and cell fate choices of retinal progenitors. Neural Dev. 2009;4(1):32.

126. Andersson ER, Sandberg R, Lendahl U. Notch signaling: Simplicity in design,

versatility in function. Vol. 138, Development. Development; 2011. p. 3593–612.

127. Olmeda D, Castel S, Vilaró S, Cano A. β-catenin regulation during the cell cycle: Implications in G2/M and apoptosis. Mol Biol Cell. 2003 Jul 1;14(7):2844–60.

128. Davidson G, Shen J, Huang YL, Su Y, Karaulanov E, Bartscherer K, et al. Cell Cycle Control of Wnt Receptor Activation. Dev Cell. 2009 Dec 15;17(6):788–99.

129. Bieler J, Cannavo R, Gustafson K, Gobet C, Gatfield D, Naef F. Robust synchronization of coupled circadian and cell cycle oscillators in single mammalian cells. Mol Syst Biol. 2014 Jul;10(7):739.

130. Hoffmann J, Symul L, Shostak A, Fischer T, Naef F, Brunner M. Non-Circadian Expression Masking Clock-Driven Weak Transcription Rhythms in U2OS Cells. Cermakian N, editor. PLoS One. 2014 Jul 9;9(7):e102238.

131. Matsuo T, Yamaguchi S, Mitsui S, Emi A, Shimoda F, Okamura H. Control mechanism of the circadian clock for timing of cell division in vivo. Science (80-). 2003 Oct 10;302(5643):255–9.

132. Karpowicz P, Zhang Y, Hogenesch JB, Emery P, Perrimon N. The Circadian Clock Gates the Intestinal Stem Cell Regenerative State. Cell Rep. 2013 Apr 25;3(4):996–1004.

133. Janich P, Pascual G, Merlos-Suárez A, Batlle E, Ripperger J, Albrecht U, et al. The circadian molecular clock creates epidermal stem cell heterogeneity. Nature. 2011 Dec 8;480(7376):209–14.

134. Jensen LD, Cao Z, Nakamura M, Yang Y, Bräutigam L, Andersson P, et al. Opposing Effects of Circadian Clock Genes Bmal1 and Period2 in Regulation of VEGF-Dependent Angiogenesis in Developing Zebrafish. Cell Rep. 2012 Aug 30;2(2):231–41.

135. Herr P, Boström J, Rullman E, Rudd SG, Vesterlund M, Lehtiö J, et al. Cell cycle profiling reveals protein oscillation, phosphorylation, and localization dynamics. Mol Cell Proteomics. 2020 Feb 12;mcp.RA120.001938.

136. Zanin E, Desai A, Poser I, Toyoda Y, Andree C, Moebius C, et al. A conserved RhoGAP limits M phase contractility and coordinates with microtubule asters to confine RhoA during Cytokinesis. Dev Cell. 2013 Sep 16;26(5):496–510.

137. Emanuele MJ, Ciccia A, Elia AEH, Elledge SJ. Proliferating cell nuclear antigen (PCNA)-associated KIAA0101/PAF15 protein is a cell cycle-regulated anaphase-promoting complex/cyclosome substrate. Proc Natl Acad Sci U S A. 2011 Jun 14;108(24):9845–50.

138. Kauffman MG, Kelly TJ. Cell cycle regulation of thymidine kinase: residues near the carboxyl terminus are essential for the specific degradation of the enzyme at mitosis. Mol Cell Biol. 1991 May;11(5):2538–46.

139. Fan H, Penman S. Regulation of protein synthesis in mammalian cells: II. Inhibition of protein synthesis at the level of initiation during mitosis. J Mol Biol. 1970 Jun 28;50(3):655–70.

140. Salazar-Roa M, Malumbres M. Fueling the Cell Division Cycle. Trends Cell Biol. 2017 Jan;27(1):69–81.

141. Horbay R, Bilyy R. Mitochondrial dynamics during cell cycling. Apoptosis. 2016 Dec 23;21(12):1327–35.

142. Mishra P, Chan DC. Mitochondrial dynamics and inheritance during cell division, development and disease. Nat Rev Mol Cell Biol. 2014 Oct 17;15(10):634–46.

143. Schaarschmidt D, Ladenburger E-M, Keller C, Knippers R. Human Mcm proteins at a replication origin during the G1 to S phase transition. Nucleic Acids Res. 2002 Oct 1;30(19):4176–85.

144. Pesavento JJ, Yang H, Kelleher NL, Mizzen CA. Certain and progressive methylation of histone H4 at lysine 20 during the cell cycle. Mol Cell Biol. 2008 Jan 1;28(1):468–86.

145. Heit R, Rattner JB, Chan GKT, Hendzel MJ. G2 histone methylation is required for the proper segregation of chromosomes. J Cell Sci. 2009 Aug 15;122(16):2957–68.

146. Katoh Y, Ikura T, Hoshikawa Y, Tashiro S, Ito T, Ohta M, et al. Methionine Adenosyltransferase II Serves as a Transcriptional Corepressor of Maf Oncoprotein. Mol Cell. 2011 Mar 4;41(5):554–66.

147. Venditto VJ, Simanek EE. Cancer therapies utilizing the camptothecins: A review of the in vivo literature. Vol. 7, Molecular Pharmaceutics. NIH Public Access; 2010. p. 307–49.

148. Li M, Liu Y. Topoisomerase I in Human Disease Pathogenesis and Treatments. Genomics Proteomics Bioinformatics. 2016 Jun;14(3):166–71.

149. Baranello L, Wojtowicz D, Cui K, Devaiah BN, Chung H-J, Chan-Salis KY, et al. RNA Polymerase II Regulates Topoisomerase 1 Activity to Favor Efficient Transcription. Cell. 2016 Apr 7;165(2):357–71.

150. Champoux JJ. DNA topoisomerases: Structure, function, and mechanism. Vol. 70, Annual Review of Biochemistry. Annu Rev Biochem; 2001. p. 369–413.

151. Leppard JB, Champoux JJ. Human DNA topoisomerase I: Relaxation, roles, and damage control. Vol. 114, Chromosoma. Springer; 2005. p. 75–85.

152. Pai DA, Engelke DR. Spatial organization of genes as a component of regulated expression. Vol. 119, Chromosoma. NIH Public Access; 2010. p. 13–25.

153. Heshmati Y, Türköz G, Harisankar A, Kharazi S, Boström J, Dolatabadi EK, et al. The chromatin-remodeling factor CHD4 is required for maintenance of childhood acute myeloid leukemia. Haematologica. 2018 Jul 3;103(7):1169–81.

154. Heshmati Y, Kharazi S, Türköz G, Chang D, Kamali Dolatabadi E, Boström J, et al. The histone chaperone NAP1L3 is required for haematopoietic stem cell maintenance and differentiation. Sci Rep. 2018 Dec 1;8(1):1–17.

155. Azimi A, Caramuta S, Seashore-Ludlow B, Boström J, Robinson JL, Edfors F, et al. Targeting CDK 2 overcomes melanoma resistance against BRAF and Hsp90 inhibitors . Mol Syst Biol. 2018 Mar;14(3).

156. Eshtad S, Mavajian Z, Rudd SG, Visnes T, Boström J, Altun M, et al. HMYH and hMTH1 cooperate for survival in mismatch repair defective T-cell acute lymphoblastic leukemia. Oncogenesis. 2016 Dec 5;5(12).

157. Rahimi Y, Goulding A, Shrestha S, Mirpuri S, Deo SK. Mechanism of copper induced fluorescence quenching of red fluorescent protein, DsRed. Biochem Biophys Res Commun. 2008 May 23;370(1):57–61.

158. Kahneman D, Klein G. Conditions for Intuitive Expertise: A Failure to Disagree. Am Psychol. 2009 Sep;64(6):515–26.

159. Jones TR, Kang IH, Wheeler DB, Lindquist RA, Papallo A, Sabatini DM, et al. CellProfiler Analyst: Data exploration and analysis software for complex image-based screens. BMC Bioinformatics. 2008 Nov 15;9(1):482.

160. Zhang HM, Chen H, Liu W, Liu H, Gong J, Wang H, et al. AnimalTFDB: A comprehensive animal transcription factor database. Nucleic Acids Res. 2012 Jan;40(D1).

161. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The Protein Kinase Complement of the Human Genome. Science (80-). 2002 Dec 6;298(5600):1912–34.

162. Cotto KC, Wagner AH, Feng Y-Y, Kiwala S, Coffman AC, Spies G, et al. DGIdb 3.0: a redesign and expansion of the drug–gene interaction database. Nucleic Acids Res. 2018 Jan 4;46(D1):D1068–73.

163. Bracken AP, Ciro M, Cocito A, Helin K. E2F target genes: Unraveling the biology. Vol. 29, Trends in Biochemical Sciences. Elsevier Current Trends; 2004. p. 409–17.